THE DESIGN OF RING-RESONATORS FOR INTEGRATED OPTICS USING SILVER ION-EXCHANGED WAVEGUIDES

A Thesis submitted to the Faculty of Engineering

of the University of Glasgow

for the degree of Doctor of Philosophy

by

ROBERT G. WALKER

January 1981

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

To my dear wife

LORNA

And God said "Let there be light"; and there was light.
And God saw that the light was good.

Genesis 1:3-4

What is man that Thou art mindful of him
and the son of man that Thou dost care for him?
Yet . . . Thou hast given him dominion over
the works of Thy hands; . . .

Psalm 8

I will give thanks to the Lord with my whole heart.

Psalm 9:1

A C K N O W L E D G E M E N T S

I wish to express my sincere gratitude to Professor J. Lamb for the provision of research facilities and materials in the Department of Electronics and Electrical Engineering (University of Glasgow). Also, I am deeply indebted to my supervisor, Dr. C.D.W. Wilkinson, for his boundless enthusiasm and unfailing helpfulness and for his criticism of the manuscript of this thesis.

Undoubtedly, the (occasionally) good results of Chapter 9 could not have been achieved without the shadow-masks made by Mr. G.W.T. Boyle, whose dedication to high-resolution photo-reduction deserves much praise and thanks.

Thanks are also due to Mr. J. Clark for his valuable assistance and advice regarding the photographic work; to Mr. K. Piechowiak who did the glass cutting and polishing; to Mrs. L. Hobbs for help in the clean-room, and to Mr. R. Hutchins for varied friendly advice and assistance.

To Dr. G. Stewart thanks are due, for many helpful discussions on ion-exchange theory. Also a debt must be acknowledged to Dr. P. Jaussaud and his colleagues at the University of Grenoble, to whom is due the basic idea which ultimately resulted in the solution of one of the central mysteries of this work (see the discussion of Chapter 8).

I must also thank Miss J. Farquhar for typing the thesis so skilfully and last, but not least, my wife for all her support.

The work was funded by the Science Research Council of the United Kingdom of Great Britain.

S U M M A R Y

This thesis is concerned with the design and ultimately, the fabrication of a ring-resonator filter for integrated optics. The resonant structure of such a filter consists of a closed, circular loop of stripe (2-D) optical waveguide which resonates whenever the circumference is an integer number of guide-wavelengths. It thus possesses a "comb" frequency response.

In order to construct the required conceptual framework, a fairly brief treatment of dielectric waveguide theory is included. This is followed by a more detailed consideration of the elements which constitute a ring-resonator: the directional coupler (used for coupling light into the ring), the curved dielectric waveguide and the resonator itself. The optimum input coupling is derived in terms of the ring losses and various theoretical frequency-response curves are given.

A major part of the thesis, and almost all of the experimental work, is devoted to the study of stripe-waveguides made by silver/sodium ion-exchange in soda-lime glass, made by immersing the glass in a bath of molten silver nitrate for a period of time. This is the type of guide which was chosen as a basis for the ring-resonator design work; its lateral refractive index difference is such as to offer a suitable compromise to the conflicting requirements of (i) low bending loss and (ii) strong directional coupling.

In order that some non-empirical design work might be possible (of the directional coupler in particular) an accurate mathematical model of the ion-exchanged waveguide is required. Accordingly, some space is given to the theory of ion-exchange and the resulting model. This takes the form of two computer programs: one to solve the non-linear partial

differential equation which describes the ion-exchange process, in two dimensions, using a finite difference technique; the other to solve the wave equation in a two-dimensional region of arbitrary, varying refractive index, using a variational technique. The substantial accuracy of these numerical solutions is demonstrated.

As essential background, and also to determine the physical parameters of the ion-exchange system, a study of slab (1-D) ion-exchanged waveguides was undertaken, comparing computed and measured guided modes. It is concluded that the computed diffusion profile differs from the real profile in certain respects; this is caused by factors for which the ion-exchange theory does not account.

Stripe waveguides are defined by coating the glass with a diffusion mask in which slit-like apertures are opened prior to diffusion. The aperture width is accurately measured by utilising Fraunhofer diffraction, and the waveguide effective refractive indices by prism-to-guide coupling.

When the diffusion mask is of aluminium, metallic silver is reduced along each edge, causing high attenuation of the guided power and other strong deviations from the computed theory. When the aluminium diffusion mask is converted to aluminium oxide, by anodisation, these effects do not occur; the anodised mask is capable of producing waveguides which behave as predicted by the mathematical model, and have low loss.

Experimental directional couplers have been made, with transfer lengths which are generally in good agreement with the computed values. Low-loss waveguide bends with radii less than 300 microns are reported, and 2 micron wide stripe waveguides with a linear attenuation of less than 3 dB/cm have been made.

Finally, design parameters for a ring-resonator are suggested, but none has yet been made.

CONTENTS

A P P E N D I C E S

CHAPTER 1

INTRODUCTION

Much research effort is currently engaged in the field of optical communications. The aim of this enterprise is to transmit data over large distances by means of modulated light waves carried by optical, dielectric waveguides. The waveguides, which are glass fibres of (usually) circular cross-section, can now be made with very low attenuation factors; figures of less than 2 dB/km are typical and less than 0.2 dB/km is now possible. Among the advantages of optical communications systems are immunity from noise due to electromagnetic or inductive pick-up, and a very large bandwidth, allowing very high digital data-rates.

In a large optical communications system, fairly complex optical signal-processing subsystems will be required (e.g. repeater stations). It is for these that the technology of "Integrated Optics" is being developed.

The integrated optics approach is a natural one because of the extremely small size of many of the optical components. Mono-mode wave-guiding structures cannot be much larger in cross-section than the wave-length of the light used - i.e. a few microns. Similarly any resonant structures used for filtering purposes should be as small as possible; otherwise the resonant modes will be so close together as to make their use in frequency selection difficult.

Because of these small sizes and the consequent difficulty of externally linking the devices together, as much as possible of the complete subsystem must be fabricated simultaneously on a single substrate.

The components are semi-conductor lasers, modulators, filters, couplers, detectors etc. and the inter-connections are lengths of mono-mode dielectric waveguide. The fabrication techniques are photo-lithographic, like those of integrated electronic circuits.

In this thesis, the device under consideration is an optical filter of the cavity-resonator type. The cavity is a closed, circular loop of optical waveguide and hence is called a Ring-Resonator. A resonant mode occurs whenever the circumference is an integer number of guide wavelengths; thus the resonator has a "comb" frequency response (see Chapter 3).

Apart from the pill-box resonator (an open, circular, dielectric cavity which behaves similarly to the ring resonator, possessing only circumferential resonant modes) the ring-resonator is perhaps the only possible planar, resonant structure which may be seriously considered as an integrated optics filter. A different, alternative approach makes use of Bragg reflection from a corrugated region of planar waveguide. The reflectivity peak of such a "grating" is reported to have a bandwidth of from 0.1 to 10 nm[64] (0.2 nm corresponds to 150 GHz at the He-Ne laser wavelength of 0.6328 μm) but there is only one main peak. A feasible ring-resonator (see Chapters 3 and 9) may have a bandwidth of only 4.5 GHz, but the comb frequency response (the peaks are 63 GHz apart for a ring of radius 0.5 mm) may prove troublesome in certain situations. Thus both approaches will probably have their particular spheres of usefulness, e.g. frequency division de-multiplexing or the resonator for an integrated laser.

This thesis, which is concerned with work leading to the design of a ring-resonator is divided into three unequal parts.

Part I, comprising Chapters 2 and 3, considers the basic theory of dielectric waveguides in general, and of the elements which constitute a ring-resonator filter in particular. These are the directional coupler, which is used to couple light to and from the ring itself, and the stripe-waveguide bend which is essentially lossy. Also the optimum input coupling is derived in terms of the ring loss - an important result.

Part II, comprising Chapters 4 and 5, is concerned with deriving a computer-based mathematical model for waveguides made by silver/sodium ion-exchange in glass. These guides, which were chosen to be the basis of the ring-resonator, are worthy of study in their own right but are not susceptible to simple, approximate analyses.

Part III, comprising Chapters 6 to 9, covers the experimental work: fabrication and measurement techniques (Chapter 6), slab and stripe waveguides (Chapters 7 and 8) and finally results for curved and straight guide attenuation and for directional couplers, culminating in a suggested ring-resonator design (Chapter 9).

A working ring-resonator has not yet been made as part of this work; however during the preparation of this thesis one was made by Haavisto and Pajer of the Northrop Corporation U.S.A.[65] (unpublished). This resonator is very different from that envisaged here, being 9 cm in diameter (too large for integrated optics), but it shows the comb-response quite clearly.

P A R T I

CHAPTER 2

INTRODUCTION TO DIELECTRIC WAVEGUIDES

The purpose of this chapter is to discuss the principles and general features of dielectric waveguides, with a view to introducing concepts and defining terminology. In constructing the necessary conceptual framework, some re-exploration of well established theory is inevitable.

A full treatment of the material of this chapter can be found in references (1) - (4).

2.1 1-Dimensional Guidance - the Slab Waveguide

The slab waveguide is the simplest form of dielectric waveguide. It consists of a layer of homogeneous dielectric material, of refractive index n_1, sandwiched between a superstrate and a substrate whose refractive indices, n_0 and n_2, are both less than n_1.

i.e. $n_1 > n_0$ or n_2 (written hereinafter $n_{0,2}$)

The slab lies in the x-z plane and supports a guided wave propagating in the z-direction.

This propagation is described by the wave equation derived in Appendix 1. If the guided wave solutions are taken to be in the form:

$$\psi(y, z, t) = \psi(y)e^{j(\omega t - \beta z)}$$

the wave equation reduces to the Helmholtz form:

$$\frac{\partial^2 \psi}{\partial y^2} + (k_i^2 - \beta^2)\psi = 0 \qquad (2.1.1)$$

where ψ represents any of the $\underline{E}$ or $\underline{H}$ field components

ω is the angular frequency

β is the propagation constant

k_i is the wave number of medium i (i = 0, 1 or 2).

Also $k_i = n_i k_0$, k_0 being the free-space wave number

($k_0 = \frac{2\pi}{\lambda_0}$ where λ_0 is the free space wavelength).

β can be normalised with respect to k_0, giving the Effective Refractive Index of the guided mode (n_e).

$$n_e = \frac{\beta}{k_0}$$

Polarisation

Only two linear polarisations need be considered;

a) T.E. (Transverse Electric) $\underline{E} = E_x$; $E_y = E_z = 0$

H_y and H_z are non-zero.

b) T.M. (Transverse Magnetic) $\underline{H} = H_x$; $H_y = H_z = 0$

E_y and E_z are non-zero.

T.E. modes are of greatest interest in this work.

Solution of the Wave Equation

If $\psi = E_x$ (T.E. modes) the general solution to 2.1.1 is:

$$E_x(y) = E_0[Ae^{j\delta_i y} + e^{-j\delta_i y}] \qquad (2.1.2)$$

where E_0 gives the field amplitude and phase reference

A is a complex constant

and:

$$\delta_i = \sqrt{k_i^2 - \beta^2} = k_0\sqrt{n_i^2 - n_e^2} \qquad (2.1.3)$$

2.1.2 may be interpreted as a superposition of two y-directed plane waves whose wave number, δ_i, is the transverse component of vector $\underline{k}_i$:

$$\underline{k}_i = (\pm \delta_i, \beta) \quad \text{where} \quad |\underline{k}_i| = k_i$$

If the solution is completed by including the factor $e^{j(\omega t - \beta z)}$, each plane wave is of form:

$$E(y, z, t) \propto e^{-j\underline{k}_i \cdot \underline{r}} \cdot e^{j\omega t} \qquad (2.1.4)$$

where $\underline{r}$ is the vector (y, z).

These plane waves, represented by rays in Fig. 2.1, are confined within the slab by successive total-internal-reflections from the two dielectric interfaces. The ray direction is given by vector $\underline{k}_1$, so the incident angle, θ, is such that:

$$\beta = k_1 \sin\theta \quad \text{and} \quad \delta_1 = k_1 \cos\theta \qquad (2.1.5)$$

If a guided wave exists within the slab region, δ_1 must be real. Hence, from 2.1.3:

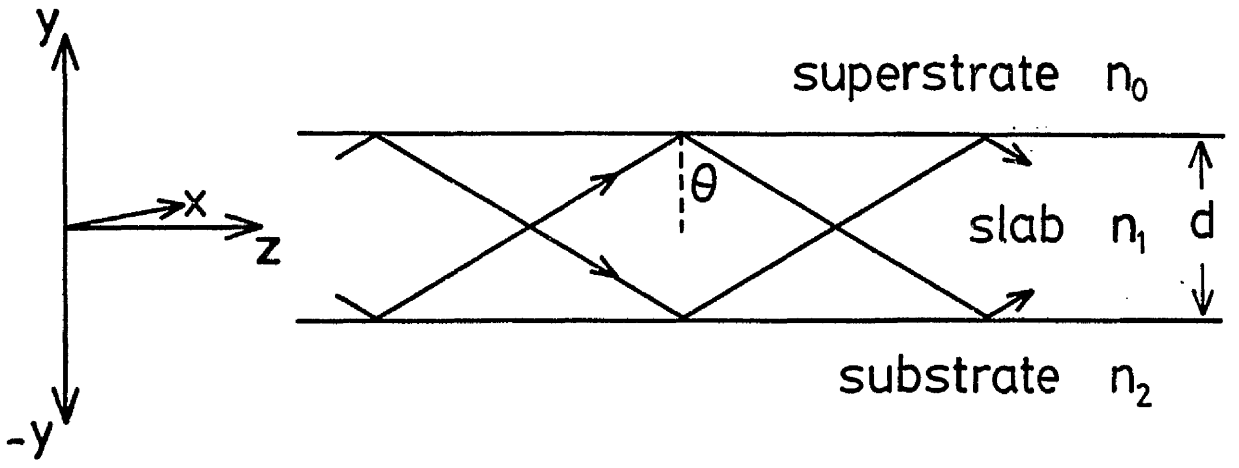
$$n_e < n_1$$

Similarly, if the energy is confined to the slab region, radiation fields may not exist in the substrate and superstrate. Therefore $\delta_{0,2}$ must be purely imaginary implying:

$$n_e > n_{0,2}$$

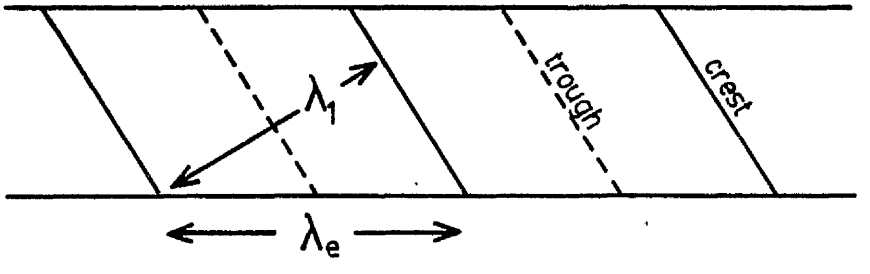
Thus, in the substrate and superstrate the guided fields are described by real exponentials, decaying away from the boundaries. These are called "evanescent fields".

Fig. 2·1



The step-index, slab waveguide. The rays represent plane waves.

Fig. 2·2



The phase fronts of one plane wave.

i.e. $$E_{x_{0,2}}(y) \propto e^{\pm \eta_{0,2} y}$$

where we define $$\eta_{0,2} = j\delta_{0,2} = k_0\sqrt{n_e^2 - n_{0,2}^2}$$

and $$k_y = \delta_1 = k_0\sqrt{n_1^2 - n_e^2} \qquad (2.1.6)$$

$\eta_{0,2}$ may be called the "evanescent decay constant".

These evanescent fields are characteristic of total internal reflection, are bound to the dielectric interface and carry energy down the guide only. The evanescent field components $E_{x_{0,2}}$ and $H_{z_{0,2}}$ are in time phase quadrature.

Cut-Off

Cut-off is defined by the condition:

$$n_e \leq n_s \qquad (2.1.7)$$

where n_s is the greater of n_0 or n_2.

If a mode is cut-off, then η_s is imaginary, and the field in region s is not evanescent but radiates away from the guide. The cut-off condition corresponds to loss of total-internal-reflection as θ falls below the critical angle θ_c.

This may be re-stated in terms of phase velocities.

The phase velocity of a guided wave, $v_e = \frac{c}{n_e}$ (where c is the velocity of light in free space), is always greater than that of the slab material, $v_1 = \frac{c}{n_1}$. This apparent anomaly may be understood by considering an effective guide wavelength, λ_e, which is greater than λ_1 (see Fig. 2.2).

An evanescent field is always "slow" however

i.e. $$v_e < v_{0,2} \quad : \quad v_{0,2} = \frac{c}{n_{0,2}}$$

If ever v_e rises above $v_{0,2}$, the velocity limit $\frac{c}{n_{0,2}}$ is exceeded unless the field breaks away from the boundary and radiates freely.

This gives a completely general definition of cut-off. It is independent of the expression for η_s which may differ in different waveguiding systems.

The Eigenvalue Equation

Within the slab, constructive interference of the plane waves of 2.1.2 is required for a guided wave to exist. The condition for this is that the total phase-shift per round-trip (two traversals of the slab plus two reflections) be an integer multiple of 2π radians. Thus:

$$2k_y d + \Phi_0 + \Phi_2 = 2m\pi \qquad (2.1.8)$$

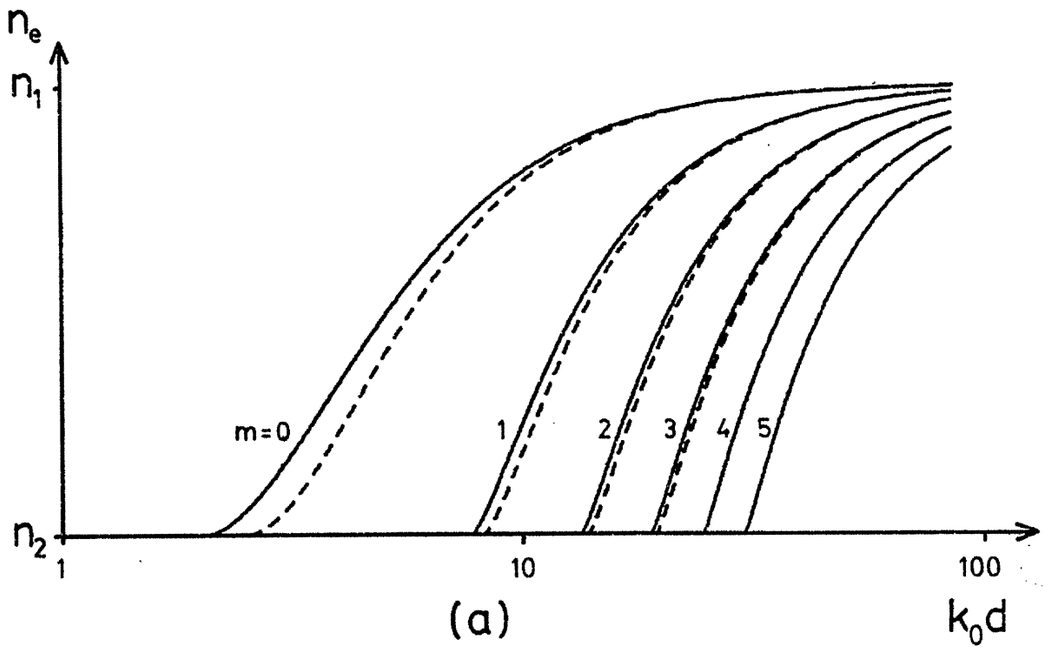
where d is the slab thickness.

m is an integer (0, 1, 2 . . .) giving the number of zeros in the standing wave pattern $E_{x_1}(y)$. m is the guided mode order.

Φ_0 and Φ_2 are the phase-shifts due to total internal reflection. They are given by Fresnel's equations or by matching tangential field components across each boundary to obtain the factor 'A' of 2.1.2.

a) T.E. modes: $\Phi_{0,2} = -2 \tan^{-1}\left[\frac{\eta_{0,2}}{k_y}\right]$

b) T.M. modes: $\Phi_{0,2} = -2 \tan^{-1}\left[\frac{n_1^2}{n_{0,2}^2} \cdot \frac{\eta_{0,2}}{k_y}\right]$ (2.1.9)



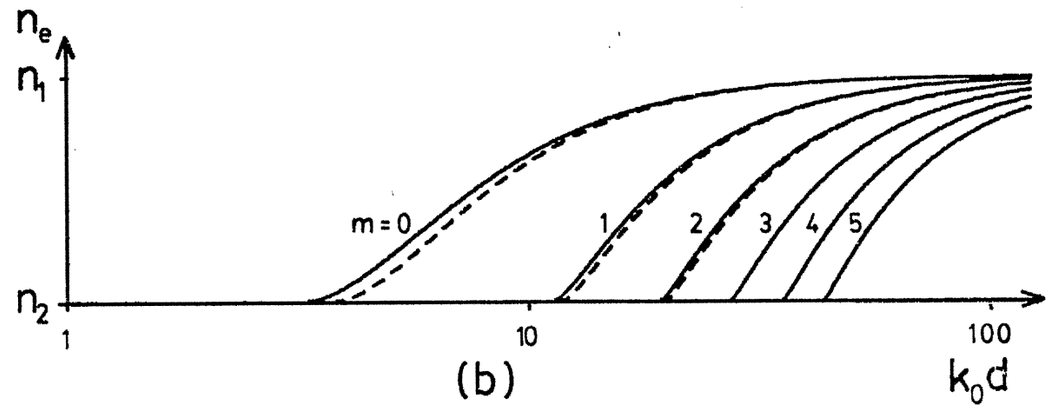


Fig. 2·3 Dispersion curves of the step-index, slab waveguide.

$n_0 = 1{\cdot}0$; $n_2 = 1{\cdot}5$; $n_1 =$ (a) 1·6 : (b) 1·55

T.E. ————

T.M. - - - - - - -

Within the slab region, the fields vary as $\cos(k_y y + P)$ if m is even (symmetric modes) and as $\sin(k_y y + P)$ if m is odd (anti-symmetric modes). See Fig. 2.7.

P is some phase constant which varies with mode order.

Waveguide Properties

The eigenvalue equation 2.1.8 is solved for two practical waveguides in Fig. 2.3. These have a glass substrate ($n_2 = 1.5$), glass slab ($n_1 = 1.6$ or 1.55) and air superstrate ($n_0 = 1$).

n_e is plotted against $k_0 d$ (log. scale) for the first 6 modes.

These graphs reveal how n_e, the mode spacings and the number of guided modes vary with k_0 or d and index-difference $(n_1 - n_2)$. Both T.E. and T.M. modes are shown.

If $(n_e - n_2)$ is small, then the evanescent field will be extensive (small n_2) leading to weak confinement of the wave. Such weakly guided modes lose more power when perturbed than well-guided modes.

It is frequently required that a waveguide support only one guided mode. Clearly, this will result in weak guidance unless the index difference $(n_1 - n_2)$ is large. However, a small index difference allows the single mode guide to be larger, thus relaxing the fabrication tolerances somewhat.

2.2 The Graded Index Slab Waveguide

For waveguiding to be possible it is necessary only that a sufficiently broad, isolated refractive index maximum exists in the transverse direction. In general, the refractive index may be a smooth function of the transverse coordinate (y). The step-index guide of 2.1 is a special case in which the refractive index profile, n(y), is a step function.

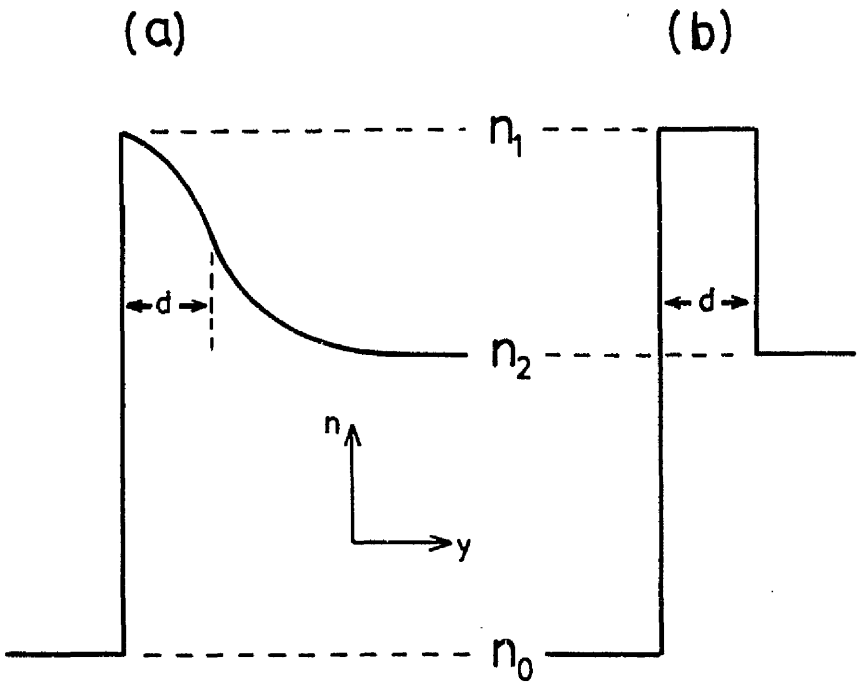


Fig. 2·4 Refractive Index Profiles:
a) Graded
b) Step

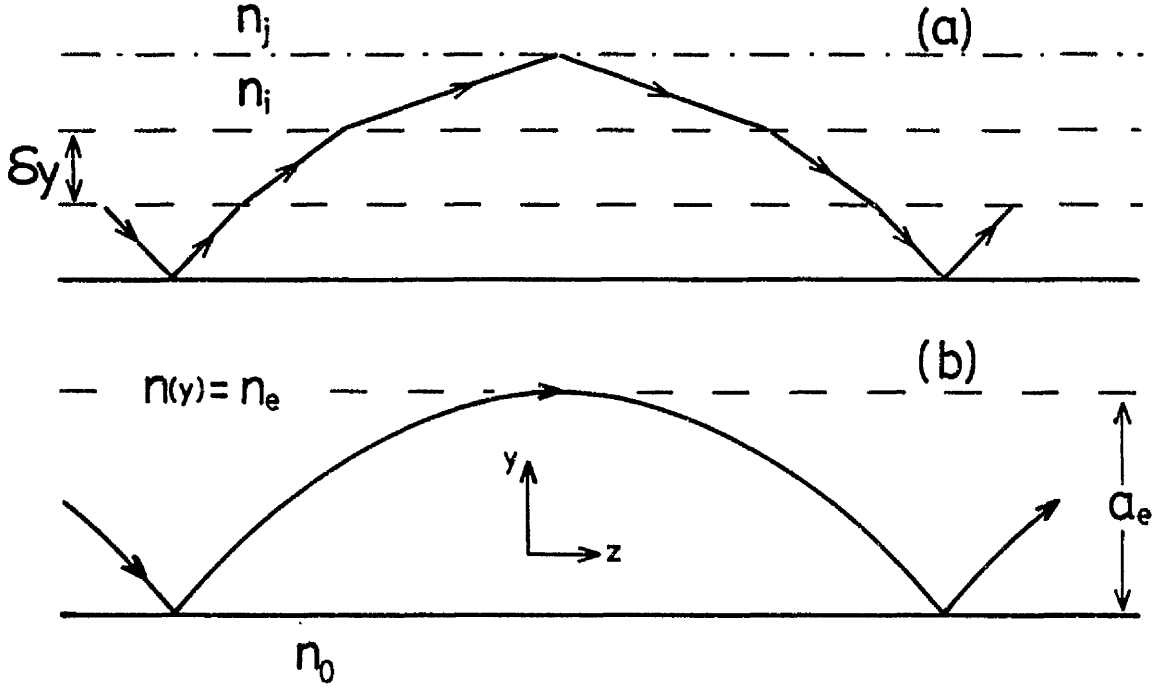


Fig. 2·5 Rays in the graded-index slab waveguide.
(a) The Layer approximation
(b) The Limit as $\delta y \to 0$

Where the waveguide fabrication involves material modification processes (e.g. diffusion of impurities) a graded refractive index profile will result. However, commonly one of the two waveguide boundaries is a step between the surface refractive index and that of air. (See Fig. 2.4). Since this is true of the ion-exchanged guides studied in this work, only this type of profile will be considered here.

The W.K.B. Method

An important method of analysis for graded index, slab waveguides is based upon the W.K.B. (Wentzel, Kramers, Brillouin) approximation which was first formulated in the field of quantum mechanics.[5],[6]

Though approximate, the W.K.B. method is found to predict the dispersion characteristics of graded index slab waveguides with considerable accuracy.[7]

The method may be developed, using the concepts of section 2.1. $n(y)$ is approximated by a series of index steps of width δy and magnitude $\delta n(y)$. The ray, which propagates through the resulting layered structure [as in Fig. 2.5 (a)], represents one of the two possible plane wave solutions in each of the constant-index layers. It is refracted into successively greater angles of incidence until eventually, if the initial conditions are favourable, a total reflection occurs. This is the only reflection considered; partial reflection from the other layer-to-layer boundaries is neglected.

In the limit as δn and δy go to zero the critical angle for total reflection is $\pi/2$. Thus, the ray must describe a smooth curve at the turning point, $y = a_e$. Since the ray is parallel to the

z-axis there, $\underline{k} = \beta$ and $n(a_e) = n_e$. No other turning point is possible since β must be a constant.

The Turning-Point Phase Shift

The above-mentioned total-reflection occurs at the boundary between two layers of index, n_i and n_j where $n_i > n_e$ and $n_j < n_e$.

We define $n_i^2 = n_e^2 + \delta(n_i^2)$; $n_j^2 = n_e^2 - \delta(n_j^2)$

where $\dfrac{\delta(n_i^2)}{\delta(n_j^2)} \longrightarrow 1$ as $\delta(n_i^2)$ and $\delta(n_j^2) \longrightarrow 0$.

The phase-shift is: $\Phi = -2 \tan^{-1} \xi^2 \sqrt{\dfrac{n_e^2 - n_j^2}{n_i^2 - n_e^2}}$

where $\xi = 1$ for T.E. modes

and $\xi = \dfrac{n_i}{n_j}$ for T.M. modes

Thus, for both polarisations:

$$\Phi = -\frac{\pi}{2} \text{ in the limit as } \delta(n_{ij}^2) \longrightarrow 0.$$

The Eigenvalue Equation

In any planar region of constant refractive index ($n > n_e$), the field amplitude, E(y), is given by the superposition of two y-directed plane waves (see 2.1.1).

In the layer approximation of the graded index slab, each layer has such a solution. Since we assume only a single, total reflection at $y = a_e$, one of each plane-wave pair must originate at the turning-point. A standing wave will result from the interference only if the total phase shift per round-trip is a multiple of 2π.

The phase-shift due to a traversal of one layer is $k_y \cdot \delta_y$

$$k_y \cdot \delta_y = k_0 \sqrt{n^2(y) - n_e^2} \cdot \delta_y \quad .$$

Collapsing the layered structure to the continuous, graded profile (i.e. $\delta_y \rightarrow 0$), the total phase shift between the surface ($y = 0$) and the turning point ($y = a_e$) is Φ_T:

$$\Phi_T = k_0 \int_0^{a_e} \sqrt{n^2(y) - n_e^2}\, dy \qquad (2.2.1)$$

Including the turning point and surface reflection phase-shift, we obtain the eigenvalue equation.

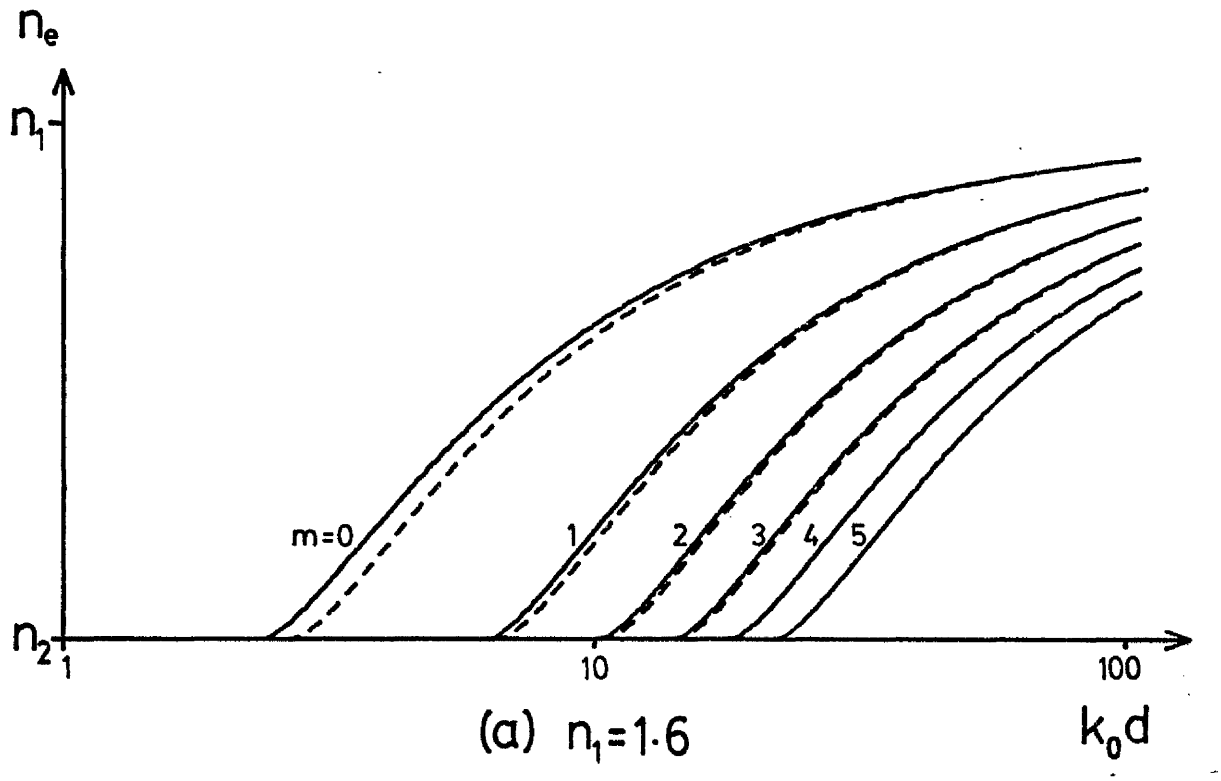
$$2k_0 \int_0^{a_e} \sqrt{n^2(y) - n_e^2}\, dy \; + \; \Phi_0 = 2m\pi + \frac{\pi}{2} \qquad (2.2.2)$$

m and Φ_0 are as defined in section 2.1.

This equation can be solved by digital computer, the resulting dispersion curves being shown in Fig. 2.6. The refractive index profile used is that of an ion-exchanged waveguide (Chapters 3 - 5). The parameter α (see Fig. 5.2) is 0.8. To facilitate comparison with the step-index slab waveguide, the same values for n_0, n_1 and n_2 as in Fig. 2.3 have been used. Here n_1 is the surface refractive index and n_2 is the limiting substrate value.

The depth d of the waveguide is (arbitrarily) defined as that for which $n(d) = \frac{n_1 + n_2}{2}$.

Comparison with Fig. 2.3 shows that graded index slabs have a smaller mode-spacing and a smaller difference between T.E. and T.M. modes than the nearest equivalent step index guides. Also the



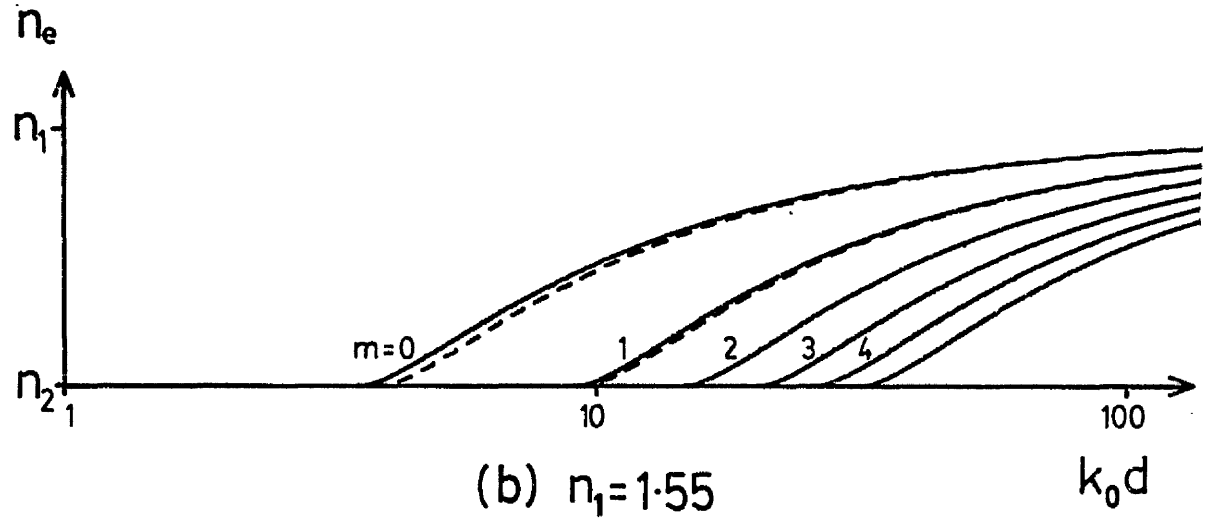


Fig. 2·6 Dispersion curves of a graded-index, slab waveguide.

n_0 = 1·0 n_2 = 1·5

T.E. ————

T.M. - - - - - -

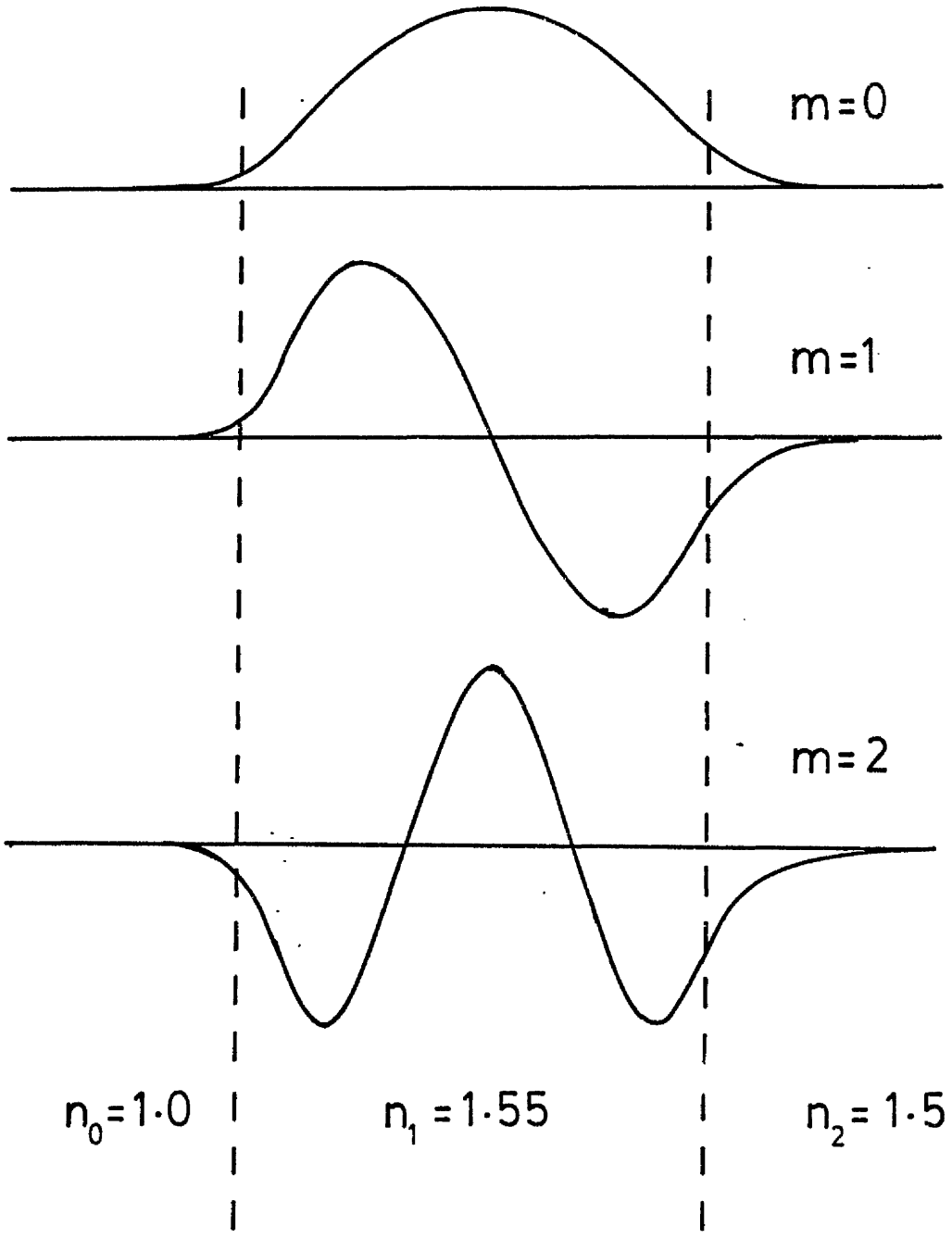


Fig. 2·7 Field amplitude distributions of a step-index slab waveguide.

dispersion curve is more nearly linear, tending only slowly towards the limiting value of n_1.

W.K.B. Errors

The errors of the W.K.B. approximation arise due to the neglect of continuous reflection.[5]

This reflection is greatest near the turning point ($y = a_e$) with the result that constructions of the field amplitude functions using the W.K.B. method become infinite at $y = a_e$.[8] However the contribution to the integral Φ_T from that region is small and the eigenvalues are accurately predicted.[7]

Additional error is incurred by T.M. modes because the exact wave equation contains the term:

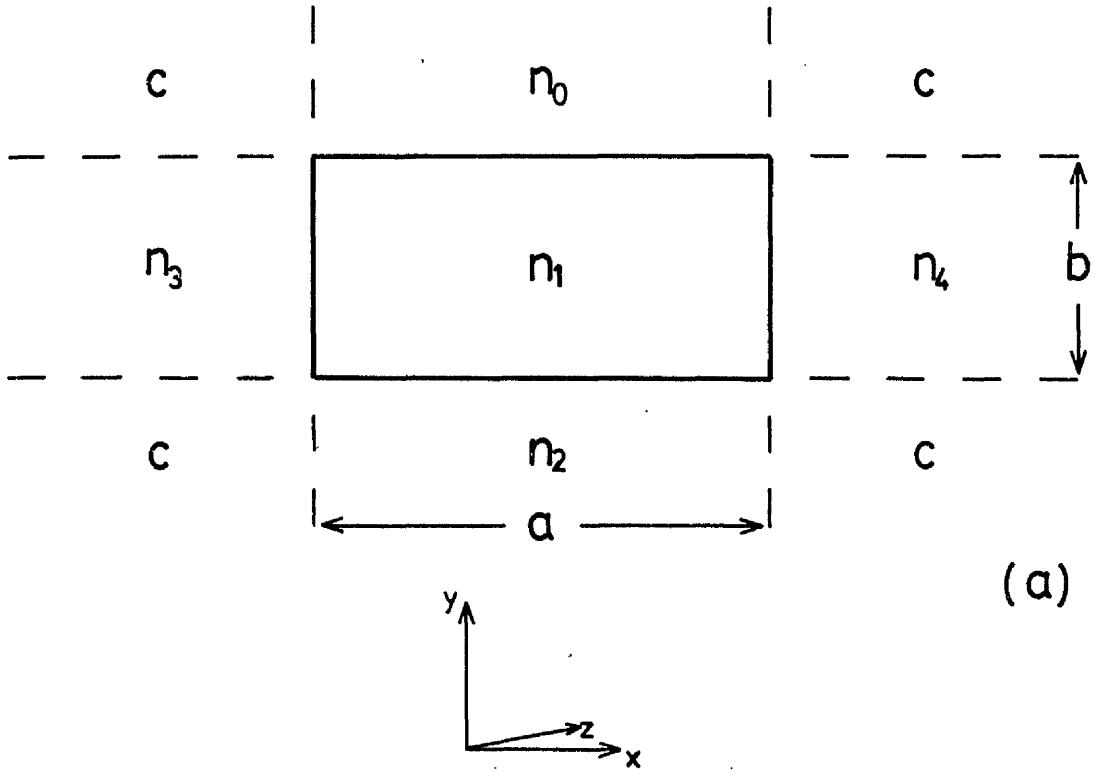
$$\text{grad}(\underline{E}.\text{grad}(\ln \varepsilon)) \qquad \text{(See Appendix 1)}$$

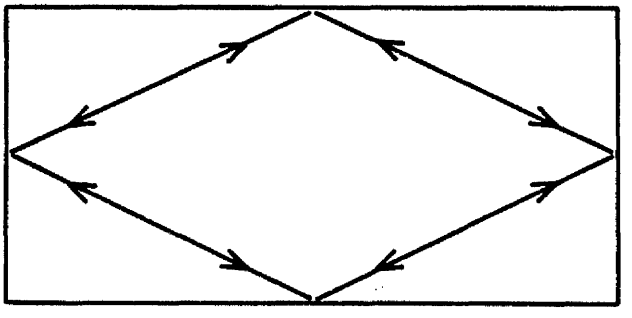
where $\varepsilon = \varepsilon_0 n^2$ - the medium permittivity.

This term is usually neglected. It is zero-valued for T.E. modes in all slab waveguides, but is non-zero if n(y) has a finite gradient in the direction of the electric field vector - as for T.M. modes in graded index slabs. However this term is negligible if the refractive index varies sufficiently slowly, as is frequently the case - see Appendix 1.

2.3 Two-Dimensional Guidance: the Stripe Dielectric Waveguide

The stripe waveguide can be developed from the slab waveguide by adding confining boundaries in the x-direction. However, in general, the waveguide cross-sectional geometry is arbitrary and need be no more than an isolated refractive index maximum in the x-y plane.





(b)

Fig. 2·8 The Rectangular Stripe Waveguide

(a) Geometry

(b) Typical ray path

The Rectangular, Step-Index Guide

This consists of a rectangular core, with refractive index n_1, surrounded by media with various, lower refractive indices. All the boundaries are parallel to the z-axis.

Assuming guided wave solutions of form:

$$\psi(x, y, z, t) = \psi(x, y)e^{j(\omega t - \beta z)} \qquad (2.3.1)$$

the wave equation is:

$$\nabla^2\psi(x, y) + (k_i^2 - \beta^2)\psi(x, y) = 0 \qquad (2.3.2)$$

where ψ may be any of the field components.

If k_i is a constant, the differential equation is linear and may be solved by separation of variables.

Assuming:

$$\psi(x, y) = \psi(x).\psi(y)$$

The general solution is:

$$\psi(x, y) = A_0e^{j\underline{k}_{r_1}\cdot\underline{r}} + B_0e^{-j\underline{k}_{r_1}\cdot\underline{r}} + C_0e^{j\underline{k}_{r_2}\cdot\underline{r}} + D_0e^{-j\underline{k}_{r_2}\cdot\underline{r}} \qquad (2.3.3)$$

where vectors: $\underline{k}_{r_1} = (k_x, k_y)$

$\underline{k}_{r_2} = (k_x, -k_y)$

$\underline{r} = (x, y)$

and $k_x^2 + k_y^2 = k_i^2 - \beta^2$

Clearly, $\underline{k}_i = (k_x, k_y, \beta)$ may be regarded as a vector giving the direction of a ray in medium i. Solution 2.3.3 is the superposition of four plane waves propagating in the x-y plane, each represented

by a section of the ray-path shown in Fig. 2.8 (b). All four waves interfere with each other at each point in space to form a 2-D standing wave.

The solution may be re-written:

$$\psi(x, y) = (A_1 e^{jk_x x} + B_1 e^{-jk_x x}).(C_1 e^{jk_y y} + D_1 e^{-jk_y y}) \quad (2.3.4)$$

Each factor of 2.3.4 must be a 1-D standing wave, therefore the waveguide may be characterised by a pair of eigenvalue equations:

$$\text{a)} \quad 2k_x a + \Phi_a = 2n\pi$$

$$\text{b)} \quad 2k_y b + \Phi_b = 2m\pi \quad (2.3.5)$$

where a and b are the guide width and depth respectively:

n and m are the (integer) mode orders of two mutually perpendicular, transverse resonances.

Φ_a and Φ_b are the total phase shifts due to all reflections encountered during two traversals of the guide width and depth respectively.

Because the waveguide boundary changes direction at the four right-angled corners, these phase-shifts are unlikely to be simple expressions and will tend to couple equations 2.3.5.

Polarisation

If the rectangular guide is excited by means of a T.E. plane wave with no x-direction propagation component, the resulting guided mode will be composed of the four plane waves of 2.3.3. Their propagation vectors and their polarisations are, respectively:

$$\underline{k} = (\pm k_x, \pm k_y, \beta) \quad ; \quad \underline{E} = (E_x, 0, \pm E_z)$$

The original T.E. wave is not completely defined by these four plane waves and the resultant mode is not T.E. The E-field lines are curved [see Goell's computer solutions [10]] and E_z is non-zero. However, E_x is usually much greater than E_z, so such hybrid modes are referred to as "Quasi-T.E." (or Quasi-T.M. if the input beam is T.M.).

In each of the four component plane-waves $\underline{E}$ is parallel to the horizontal, and H to the vertical interfaces. Therefore, if we assume that the reflection phase shifts given by equations 2.1.9 are applicable to the rectangular guide, these phase shifts are:

$$\text{a) Horizontal Interfaces:} \quad \Phi_{0,2} = -2 \tan^{-1}\left[\frac{\eta_{0,2}}{k_y}\right]$$

$$\text{b) Vertical Interfaces:} \quad \Phi_{3,4} = -2 \tan^{-1}\left[\frac{n_1^2}{n_{3,4}^2} \cdot \frac{\xi}{k_x}\right] \qquad (2.3.6)$$

where refractive indices n_0 to n_4 are as shown in Fig. 2.8 (a).

The evanescent decay constants are:

$$\eta_{0,2} = \sqrt{\beta^2 - k_{0,2}^2 + k_x^2}$$

$$\xi_{3,4} = \sqrt{\beta^2 - k_{3,4}^2 + k_y^2} \qquad (2.3.7)$$

However, these expressions violate the general cut-off condition 2.1.7. At cut-off ($n_e = n_s$) the evanescent field in medium s should spread infinitely - i.e. $\eta_s = 0$.

But, from 2.3.7, if $n_s = n_2$ (for example)

$$\eta_2 = k_x \neq 0.$$

Equations 2.3.6 and 2.3.7 are invalid in this case. They were derived for an infinite, plane dielectric interface and cannot apply in the vicinity (i.e. within a few effective, transverse wavelengths) of the boundary direction change at the corners. To obtain the correct expressions, a rigorous matching of the fields in the regions 'C' of Fig. 2.8 (a) would be required, but this is impractical.

Nevertheless, for well guided modes, only a small fraction of the guided energy will be present in regions 'C' and the error incurred by using 2.3.6 and 2.3.7 will be small. Alternatively, since the effective, transverse wavelength is small much of the guide perimeter may be considered to be sufficiently far from the corners for these expressions to apply.

Near cut-off, however, 2.3.6 and 2.3.7 cannot be valid.

This consideration reveals one serious difficulty in accurately analysing stripe dielectric waveguides; the reflection phase shifts are not given by any simple, analytical formula - even for this simple geometry.

However, the cut-off criterion suggests that, near cut-off, the evanescent field in medium s (where n_s is the largest surrounding index) is described approximately by the constant η_s:

$$\eta_s = k_0 \sqrt{n_e^2 - n_s^2} \qquad (2.3.8)$$

Two well-known approximate solutions will be considered:

i) <u>The Method of Marcatili</u>

Marcatili's approach[9] is based upon equations 2.3.5 in which (from 2.3.6):

$$\Phi_a = \Phi_3 + \Phi_4 \quad \text{and} \quad \Phi_b = \Phi_0 + \Phi_2$$

(9) gives exact solutions of this formulation as well as approximate closed-form solutions and compares these with the more accurate computer analysis of Goell.[10] The comparison - especially of the fundamental mode - is rather poor near cut-off but good far from cut-off. This is to be expected in view of the above considerations.

ii) The Effective Index Method

This method of analysis [due to Knox and Toulios[12]] is also based upon equations 2.3.5.

When the guide is far from cut-off and is rather wider than it is deep, k_y is close to that of the slab, formed by extending the long dimension (a) to infinity. Assuming k_y to be independent of width, 2.3.5 (b) may be solved to give the slab β (β_0):

$$\beta_0^2 = \beta^2 + k_x^2$$

and β_0 is assumed to be independent of k_x.

The lateral resonance, 2.3.5 (a) becomes:

$$k_0 \sqrt{n_{e_0}^2 - n_e^2} \cdot a + \Phi_a = 2n\pi$$

where $n_{e_0} = \dfrac{\beta_0}{k_0}$

This resembles the equation of a vertical slab of width 'a' and refractive index n_{e_0}. Φ_a is chosen accordingly; appropriately, with T.M. polarisation:

$$\Phi_a = \Phi_3 + \Phi_4$$

where

$$\Phi_{3,4} = -2 \tan^{-1}\left[\frac{n_{e_0}^2}{n_{3,4}^2} \cdot \frac{\xi_{3,4}^2}{k_x^2}\right]$$

and

$$\xi_{3,4} = k_0 \sqrt{n_e^2 - n_{3,4}^2}$$

Although this choice is consistent with the cut-off condition, it has no other theoretical justification.

The Effective index method is more accurate near cut-off than Marcatili's method, as shown by Hocker and Burns.[11] This is due to the choice of Φ_a and to the introduction of some coupling between the equations 2.3.5 (k_x is made dependent upon k_y via β_0).

Non-Rectangular Geometries

Solutions of the wave equation of form 2.3.4 and 2.3.5 are only possible for rectangular waveguides. The wave equation 2.3.2 is separable in several coordinate systems and the most suitable must be selected. For example, with a circular fibre, cylindrical coordinates would be used:

$$\text{i.e.} \quad \psi(x, y) \neq \psi(x).\psi(y)$$

$$\text{but} \quad \psi(r, \theta) = \psi(r).\psi(\theta)$$

The solution will be the product of a circumferential and a radial standing wave[2].

Similarly, with a waveguide of trapezoidal or triangular cross-section, k_x and k_y would assume different values after each reflection (with $k_x^2 + k_y^2$ constant) making a solution of form 2.4.4 impossible.

The Diffused Channel Guide

The present work is entirely concerned with this type of waveguide. It is relatively easily made and possesses excellent guiding properties.

The diffused channel guide is made by the selective doping of a planar substrate by an index modifier, producing a two-dimensional graded index channel beneath the surface. At this surface there is a large index step.

If, as is common, a diffused guide is much wider than it is deep, it may be considered to be approximately rectangular and will behave in a qualitatively similar manner to the rectangular guide.

It is not feasible to extend the W.K.B. method to two dimensions; however the following may be deduced:

(1) The $n(x, y) = n_e$ contour marks the boundary between the confined and the evanescent fields. This follows from a consideration of phase velocities as in section 2.1.

(2) It follows from (1) that the effective width and depth of the guide vary with n_e. This will affect the manner in which the mode spacing changes with mode order or diffusion depth.

With diffused channel guides, the need for an accurate, numerical, general solution of the wave equation becomes acute. The Effective Index Method cannot be used with confidence unless it has been first checked against such a solution and a calibration obtained for the effective width.

2.5 Summary of the General Characteristics of Stripe Dielectric Waveguides

If the waveguide geometry is roughly rectangular i.e. it has a long and a short dimension whose axes are mutually perpendicular, the following qualitative statements may be made.

(1) A pair of distinct but not independent transverse resonances across the long dimension (width) and the short dimension (depth) may be distinguished. Associated with these resonances are what may be termed the "lateral" mode order (n) and the "slab" mode order (m) respectively. m and n give the number of field zeros of the quasi-sinusoidal standing wave pattern across the width and depth respectively.

(2) Associated with each slab mode is a full spectrum of lateral modes. These respond to width and index-difference in the same way as do slab modes to depth and index-difference. (See Figs. 2.3 and 2.6).

(3) For well guided modes the resonances are only slightly coupled; the lateral evanescent field is determined largely by the width and the lateral index difference. For modes near cut-off, the transverse resonances are strongly coupled, so that the index difference $n_e - n_i$ governs the evanescent field in medium i, irrespective of width or depth.

(4) A large index difference, $n_1 - n_i$, allows a single-mode guide to have a large n_i (well confined field). This is only achievable with a smaller index difference if the guide is multimode.

CHAPTER 3

THE RING-RESONATOR AND ITS ELEMENTS

The ring-resonator is a closed loop of stripe, optical waveguide. In practice, the optimum shape will be a circle (since curvature changes cause loss), although the resonance condition is unaffected by the shape.

In its application as a channel dropping filter, the ring-resonator must extract light of one wavelength from the input waveguide while leaving that of other wavelengths undisturbed. This can be achieved only by using directional coupling to transfer power from the input guide to the ring.

3.1 The Directional Coupler

The directional coupler consists of two parallel, dielectric waveguides whose separation is sufficiently small for their evanescent fields to overlap. As will be shown, this results in a transference of power between the waveguides. The fraction of power transferred from one guide to the other depends upon the distance, S, for which the guides run parallel before diverging. S may be termed the "interaction length".

Since the ring itself will form one branch of the coupler, the input guide must, necessarily, follow the curve of the ring for a fraction of its circumference in order to secure sufficient power transfer (see Fig. 3.1). It will be assumed that the resulting curved directional coupler behaves in essentially the same way as a straight one.

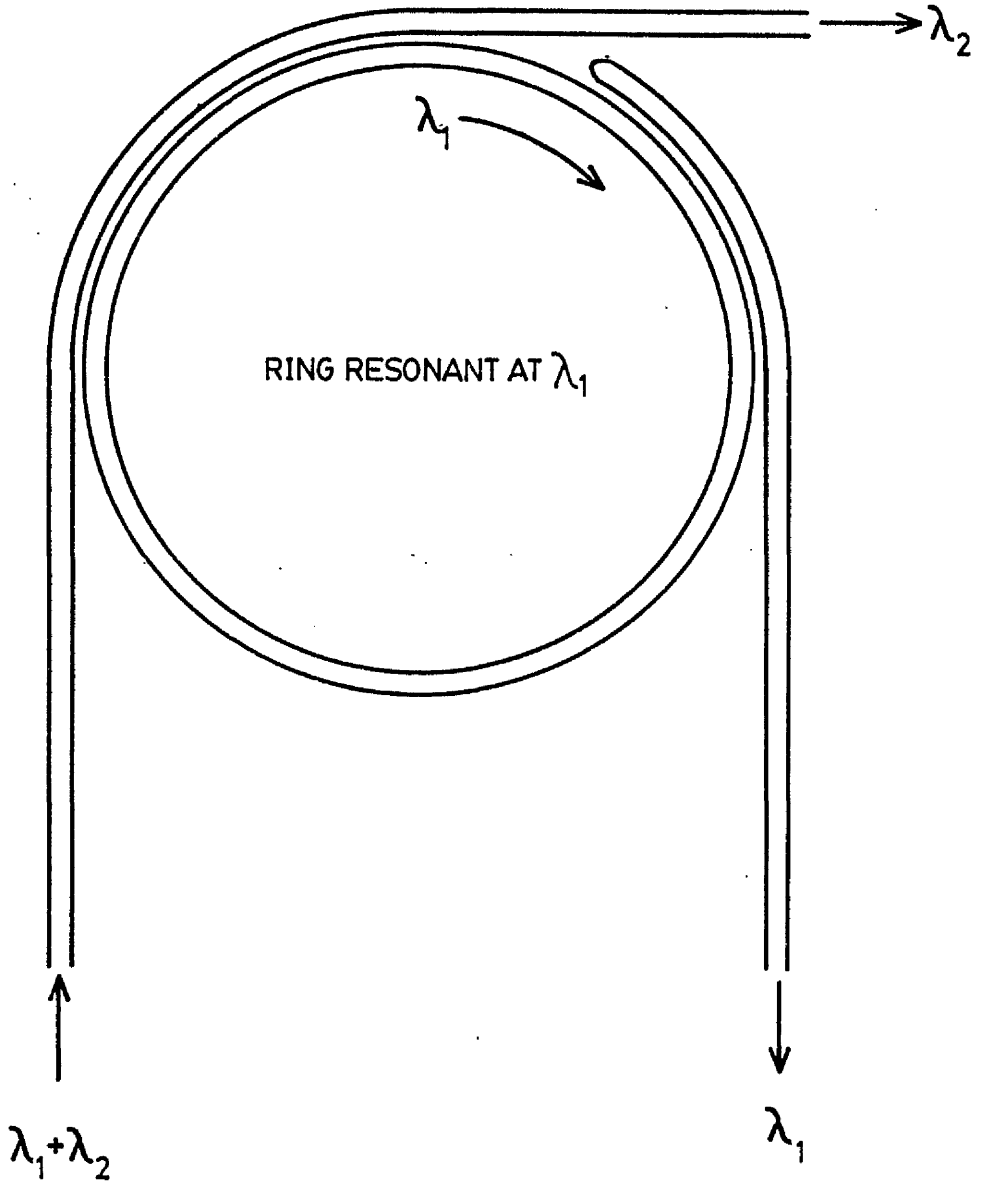


Fig. 3.1 Schematic diagram of Ring-Resonator

The coupler may be developed by considering the two identical, parallel waveguides to be initially very far apart. The wave amplitude function of each guide is ψ_0 (x,y). We consider only the fundamental mode.

If the guide separation is reduced, the guided fields will overlap and a mode of the composite guiding structure will result from their superposition. This mode will be symmetric if the guides propagate in phase, or antisymmetric, if in antiphase.

$$\text{(i) Symmetric:} \quad \psi_s(x,y) = \psi_0[(x - x_0),y] + \psi_0[(x_0 - x),y]$$

$$\text{(ii) Antisymmetric:} \quad \psi_a(x,y) = \psi_0[(x - x_0),y] - \psi_0[(x_0 - x),y] \qquad (3.1.1)$$

where the guides are centred at $\pm x_0$. See Fig. 3.2.

The symmetrical, composite, guiding structure cannot support a pure mode (i.e. one whose amplitude has no z-dependence) which is asymmetric.

These two modes have different propagation constants (β_s and β_a) which diverge as the guide spacing, c, is reduced.

When $c = \infty$, $\beta_s = \beta_a = \beta_0$

When $c = 0$, $\beta_s = \beta_{2_0}$ and $\beta_a = \beta_{2_1}$

where β_{2_0} and β_{2_1} are the lowest-order lateral modes of the double-width guide.

Thus β_s is greater than β_a and we define:

$$\beta_s = \beta_0 + \delta\beta$$
$$\beta_a = \beta_0 - \delta\beta \qquad (3.1.2)$$

Because of this divergence, β_0 and ψ_0 in 3.1.2 and 3.1.1 will not be strictly identical to those of the isolated guide. However, if $\delta\beta$

is small - as will be assumed - 3.1.1 may be used with very little error.

If all the power is initially in one guide, this asymmetric amplitude distribution can only be due to the presence, in phase and in equal strengths, of both the symmetric and the antisymmetric modes. The amplitude function $\psi_c(x,z)$ (neglecting, temporarily, the y-dependence) is their sum.

$$\psi_c(x,z) = [\psi_0(x - x_0) + \psi_0(x_0 - x)]e^{-j\beta_s z} + [\psi_0(x - x_0) - \psi_0(x_0 - x)]e^{-j\beta_a z}$$

Using 3.1.1 and rearranging, we obtain:

$$\psi_c(x,z) = \left[\psi_0(x - x_0)\,[e^{j\delta\beta.z} + e^{-j\delta\beta.z}] - \psi_0(x_0 - x)\,[e^{j\delta\beta.z} - e^{-j\delta\beta.z}]\right]e^{-j\beta_0 z}$$

whence:

$$\psi_c(x,y,z) = [\psi_1(x,y)\cos(Kz) - j\psi_2(x,y)\sin(Kz)]\,e^{-j\beta.z} \qquad (3.1.3)$$

where $\psi_1(x,y) = 2\psi_0[(x - x_0),y]$; $\psi_2(x,y) = 2\psi_0[(x_0 - x),y]$

and where $K = \delta\beta$ is the Coupling Coefficient.

Thus, the coupling phenomenon may be interpreted as the "beating" of two almost degenerate modes. The power associated with each guide varies as the square of the field amplitude.

i.e. Power associated with guide 1 $\propto \cos^2(Kz)$

Power associated with guide 2 $\propto \sin^2(Kz)$

If all the power is initially in guide 1, it will be completely transferred to guide 2 over a distance L where:

$$KL = \pi/2$$

$$\text{Therefore} \quad L = \frac{\lambda_0}{2(n_{e_s} - n_{e_a})} \qquad (3.1.4)$$

L is called the "Transfer Length".

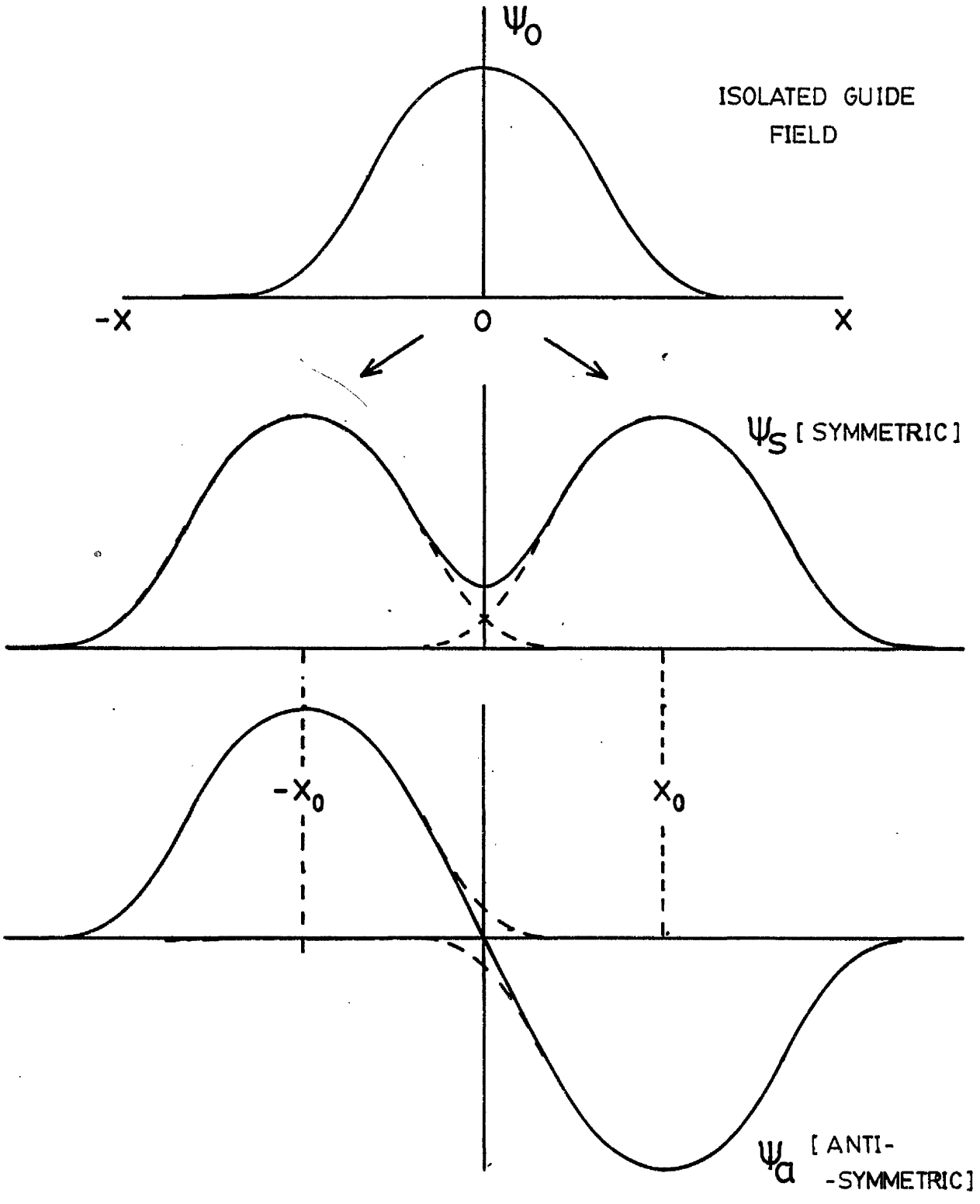


Fig. 3·2 Construction of directional coupler modes.

Evanescent Coupling

As an alternative, complementary approach to the directional coupler we consider the effect of the evanescent field of guide 1 penetrating guide 2. Within the guide, n is greater than n_e, so the evanescent field must convert to a radiation field and a guided mode be excited. The coupling coefficient is a function of that fraction of the evanescent field energy which is within guide 2. This energy fraction is proportional to the square of the field amplitude at the inner boundary of guide 2. Knowing the limits of K for $c = \infty$ and $c = 0$, it may be inferred:

$$K(c) = K_0 e^{-\varepsilon \xi c} \qquad (3.1.5)$$

where $K_0 = K(0) = \frac{1}{2}(\beta_{2_0} - \beta_{2_1})$ (see p.24)

ε is some constant

ξ is the evanescent decay constant.

Equation 3.1.5 shows the requirements for strong coupling between parallel guides:

(i) Large K_0: The guides should be narrow if the double width guide is to have a large mode separation $(\beta_{2_0} - \beta_{2_1})$.

(ii) Small ξ: The medium between the waveguides should have a refractive index which is close to n_e.

(iii) Small c: The guides should be close together.

Each mode will have its own coupling coefficient; in general K increases with mode order.

If the guides are not identical $(\beta_1 \neq \beta_2)$ then the transfer of power will not be complete for $s = L$.[9],[16],[19] With the planar fabrication techniques used in this work, this would most probably be caused by differences in the guide widths. Therefore it is undesirable for the guides to have a large width-dispersion (i.e. the rate of change of n_e with respect to width.)

3.2 The Resonator

The ring-resonator differs from other cavity resonators in that, at resonance, there is no standing wave. Propagation is unidirectional with respect to a curved propagation coordinate.

We consider a continuous wave, with amplitude ψ_0, coupled into the ring at a point 'A' (see Fig. 3.3) situated at the end of the input coupler. This wave is guided around the closed loop and arrives back at A shifted in phase by βl; β is the guide propagation constant and l is the ring circumference. The curved guide is lossy (the input coupler contributes to this loss), so the wave amplitude is reduced by a factor σ. Arriving back at A, this wave encounters and adds to another part of the wave-train which has just coupled in. The resultant wave goes round the ring and this sequence is repeated until the initial wave-front has decayed away to zero amplitude. Thus the steady-state amplitude at A is given by the following infinite sum:

$$\psi_r = \psi_0 \sum_{n=0}^{\infty} \sigma^n e^{-jn\beta l} \qquad (3.2.1)$$

ψ_r is maximised when $\beta l = 2m\pi$ (resonance)

and minimised when $\beta l = 2(m + \frac{1}{2})\pi$ (anti-resonance).

where m is the (integer) order of resonance.

The m.th resonant frequency is:

$$f_{0_m} = \frac{mc}{n_e l}$$

where c is the velocity of light in free space.

The ring-resonator thus has a 'comb' frequency-response with a constant resonant-frequency spacing:

$$\Delta f = \frac{c}{n_e l}$$

The Input Coupler at Resonance

Being a symmetrical device, the input coupler contributes to the loss of the ring. This coupling must, therefore, be optimised if the resonance is not to be damped unduly.

If the input coupler has coupling coefficient, K, and interaction length, s, the factor σ may be written:

$$\sigma = \gamma \cos(Ks)$$

where γ may be called the "ring transmission factor", and accounts for both the intrinsic ring losses and the output coupling.

If the amplitude of the wave of the beginning of the input guide is ψ_1 then, from 3.1.3:

$$\psi_0 = -j\psi_1 \sin(Ks)$$

Thus, at resonance, from 3.2.1

$$\psi_r^r = -j\psi_1 [\sin(Ks) \cdot \sum_{n=0}^{\infty} \gamma^n \cos^n(Ks)] \qquad (3.2.2)$$

Being a simple power series, the sum may be calculated immediately. At resonance:

$$\psi_r^r = -j\psi_1 \frac{\sin(Ks)}{1 - \gamma\sqrt{1 - \sin^2(Ks)}} \qquad (3.2.4)$$

For optimum coupling we require to maximise the amplitude in the ring. Thus

$$\frac{\partial \psi_r^r}{\partial [\sin(Ks)]} = 0$$

Some simple calculus yields the result:

$$\cos(Ks) = \gamma \qquad (3.2.5)$$

Thus, for optimum coupling, the power input, $\sin^2(Ks)$, must balance the power loss per circuit $(1 - \gamma^2)$.

For this optimum:

$$\psi_r^r = -j\psi_1 \frac{1}{\sqrt{1-\gamma^2}} \qquad (3.2.6)$$

3.2.6 gives the amplitude in the ring at point A (the end of the coupler). At the start of the coupled section the amplitude is:

$$\gamma\psi_r^r$$

Therefore, the amplitude in the input guide, at point A, is:

$$\psi_2^r = \psi_1 \cos(Ks) - j\gamma\psi_r^r \sin(Ks)$$

Substituting the optimum coupling, 3.2.5 and 3.2.6 we find:

$$\psi_2^r = 0$$

Thus, at resonance and for optimum coupling, the input power is totally diverted into the ring.

The Input Coupler at Anti-Resonance

At anti-resonance $(\beta l = 2(m + \tfrac{1}{2})\pi)$

$$\psi_r^a = -j\psi_1[\sin(Ks) \sum_{n=0}^{\infty} (-1)^n \gamma^n \cos^n(Ks)]$$

With optimum coupling $(\cos(Ks) = \gamma)$, this yields for the input guide and the ring, respectively:

$$\psi_2^a = \psi_1\gamma\left[1 + \frac{1-\gamma^2}{1+\gamma^2}\right]$$

$$\psi_r^a = -j\psi_1 \frac{\sqrt{1-\gamma^2}}{1+\gamma^2}$$

Thus a real, lossy ring-resonator with adequate input coupling will have a wave amplitude which is non-zero at all frequencies.

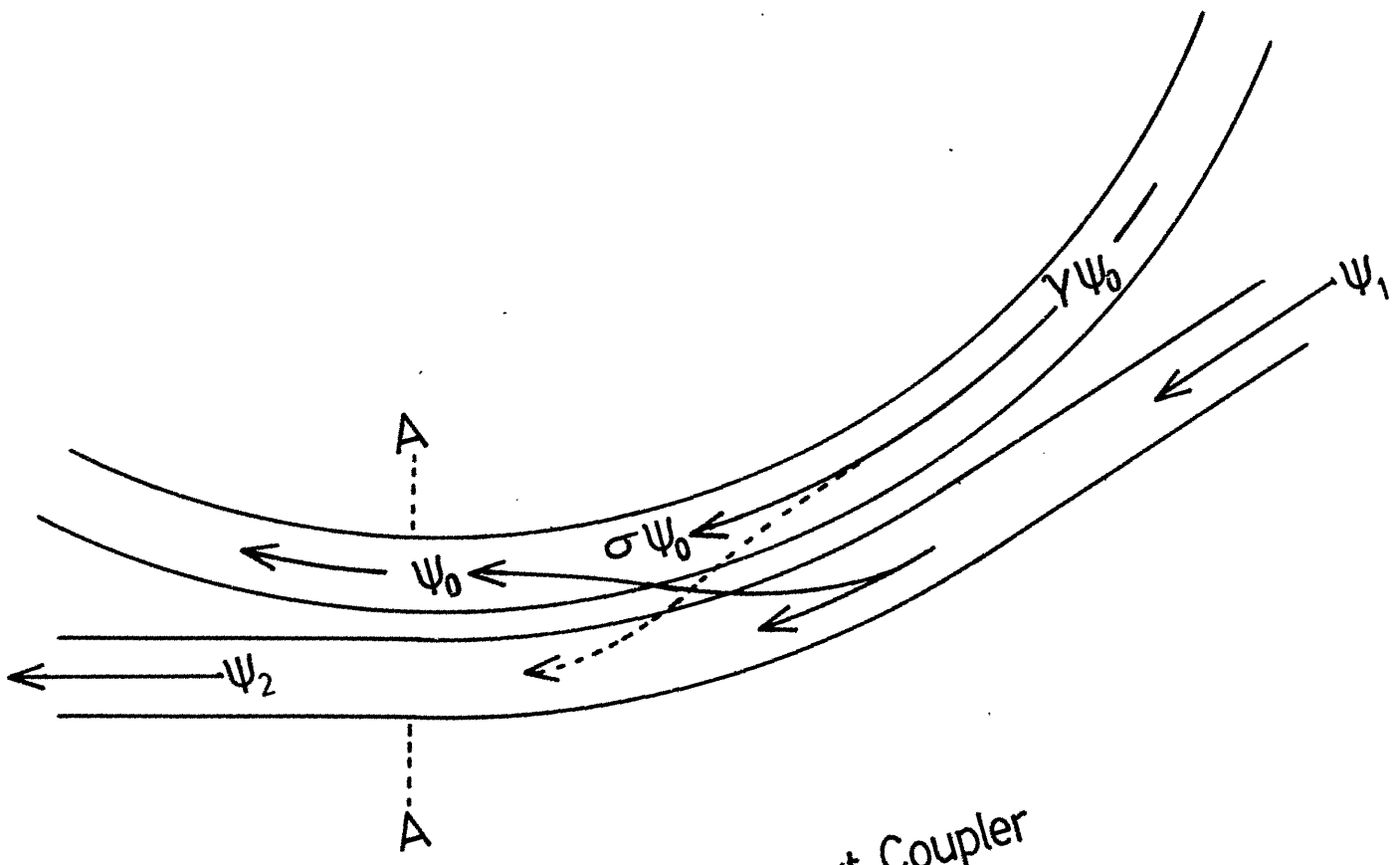


Fig. 3.3

The Ring-Resonator Input Coupler

Output Coupling

If an output coupler identical to the input coupler were placed at point A, it is clear from 3.2.6 that, at resonance, the output guide amplitude, ψ_3^r, would equal $-\psi_1$. The ring would shunt the input energy into the output guide entirely. This is the ideal for a channel dropping filter.

However, since, for optimum coupling:

$$\sigma = \cos^2(Ks)$$

this ideal can be achieved only if the ring has no intrinsic loss.

In general, output coupling should be kept as low as is possible for the purpose envisaged. Also, it may be advantageous to under-couple the input, thus sacrificing some efficiency for an improved Q-factor. [See Fig. 3.6 (a)]. Over-coupling should always be avoided as it reduces both the efficiency and the Q-factor. [See Fig. 3.6 (b)].

The Q-Factor of a Ring Resonator

The basic definition of the Q-factor is:

$$Q = \frac{\omega_0 \times \text{Energy Stored}}{\text{Mean Power Loss}} \qquad (3.2.7)$$

where ω_0 is the resonant angular frequency.
To derive the expression for the Q-factor of a ring resonator we consider a discrete packet of photons travelling one circuit of the ring. The initial energy of the packet is w_0.

Energy after one circuit $= \sigma^2 w_0$

Energy after two circuits $= \sigma^4 w_0$

On average, we may say, $w = w_0 \exp(-\alpha z)$

where z is the curved path coordinate and α is the mean attenuation constant.

From this it follows that $\alpha = \frac{\ln(\sigma^2)}{l}$

In terms of time: $w = w_0 \exp(-\alpha . v_g . t)$

v_g is the velocity of the energy packet. This is the waveguide group velocity. Stripe waveguides are by no means dispersionless.

$$\text{Mean Power Loss} = -\frac{dw}{dt}$$

$$= \alpha v_g w$$

whence:

$$Q = \frac{2\pi f_0}{\alpha v_g}$$

$$= -\frac{2\pi f_0 l}{2\ln(\sigma) v_g}$$

By definition group velocity $v_g = \frac{d\omega}{d\beta}$

$$\text{So} \quad \frac{1}{v_g} = \frac{d\beta}{d\omega}$$

$$= \frac{1}{2\pi}\frac{d}{df}(k_0 n_e)$$

Changing the variable to λ and expanding gives:

$$\frac{1}{v_g} = \frac{\lambda_0}{c}\left[\frac{n_e}{\lambda_0} - \frac{dn_e}{d\lambda}\right]$$

So Q-factor (in terms of ring radius R) is:

$$Q = -\frac{2\pi^2 R}{\ln(\sigma)}\left[\frac{n_e}{\lambda_0} - \frac{dn_e}{d\lambda}\right] \qquad (3.2.8)$$

Since $\frac{dn_e}{d\lambda}$ is itself negative, the Q-factor is seen to be increased by the waveguide dispersion. Using the technique described in Chapter 5 to compute the modes of stripe waveguides, the dispersion term $\frac{dn_e}{d\lambda}$ is found to be less than 10% of $\frac{n_e}{\lambda}$ in the area of concern - i.e. near cut-off to obtain single mode guidance. Thus, to a first approximation, this term may be neglected.

The Frequency Response

In terms of frequency, and with optimum input coupling, the amplitude of the wave in the ring may be written (from 3.2.1):

$$\psi_r = -j\psi_1\sqrt{1-\gamma^2}\left[\sum_{n=0}^{\infty}\left(\gamma^{2n}\cos\left(2n\pi\frac{f}{\Delta f}\right)\right) + j\sum_{n=0}^{\infty}\left(\gamma^{2n}\sin\left(2n\pi\frac{f}{\Delta f}\right)\right)\right]$$

This is calculated over a range of frequencies by digital computer. For each frequency the series is summed until γ^{2n} is less than 0.01. The Q-factor may also be calculated using the formula:

$$Q = \frac{f_0}{\delta f} \qquad (3.2.9)$$

where δf is the interval between the half-power points.

Fig. 3.4 shows how the frequency response (ψ_r vs. frequency) varies with γ, assuming that the input coupling is maintained at the optimum.

Fig. 3.5 shows the comb-frequency response and demonstrates how the resonance-spacing, Δf, and the Q-factor vary with the ring radius R.

Fig. 3.6 illustrates the effect of (a) undercoupling and (b) overcoupling the input. Clearly it is not particularly disadvantageous to undercouple; reducing the optimum interaction length by up to half.

The figures for the Q-factor given by formulae 3.2.8 and 3.2.9 are in good agreement, the error being about 5%.

Clearly, Q-factor values for ring resonators must be treated with caution as very poor quality resonators [Fig. 3.4(c)] nevertheless have large Q-factors. It is suggested that an "acceptance-ratio", $\frac{\psi_r^r}{\psi_r^a}$, be used in addition. With optimum coupling:

$$\frac{\psi_r^r}{\psi_r^a} = \frac{1+\gamma^2}{1-\gamma^2} \qquad (3.2.10)$$

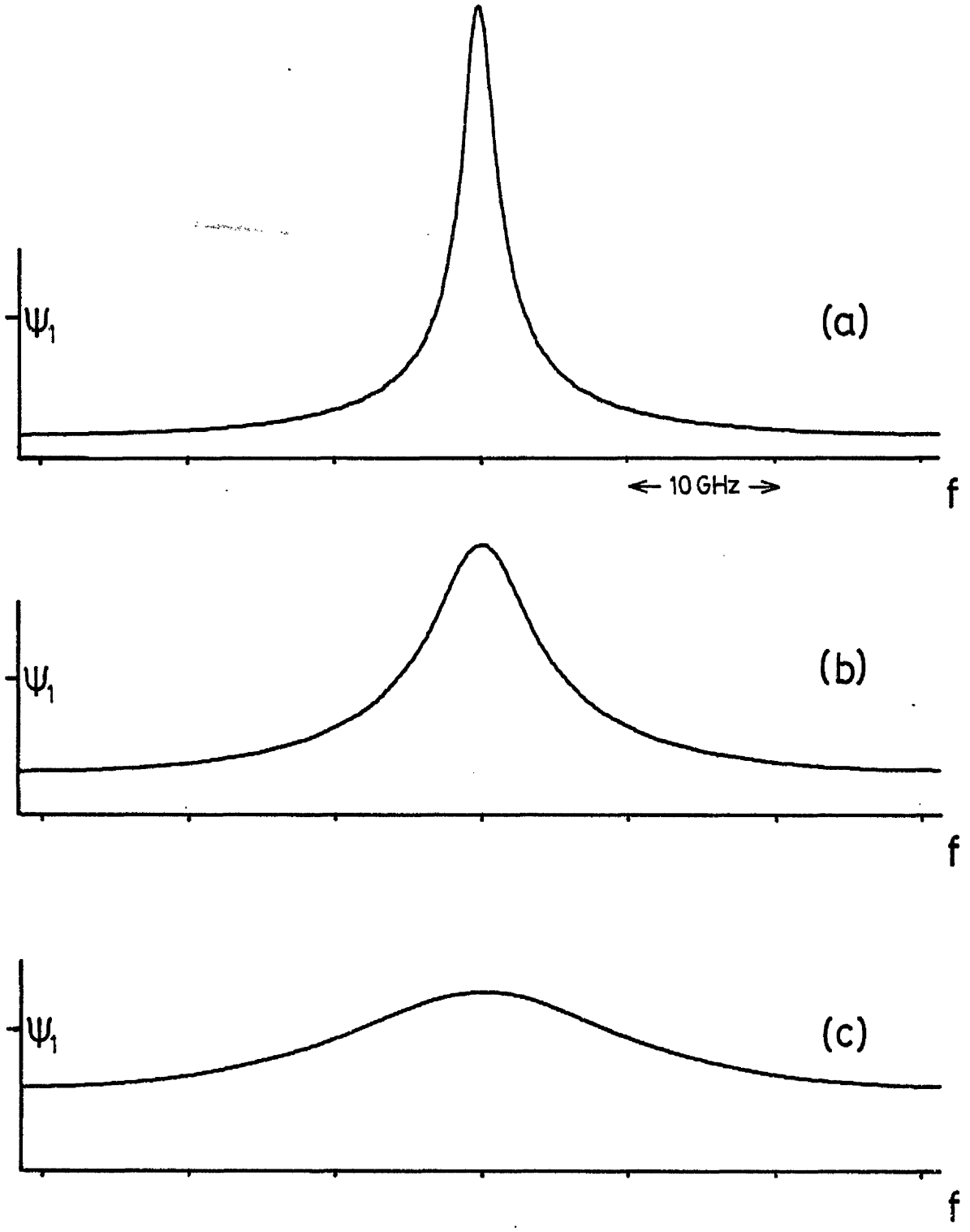


Fig. 3·4 Theoretical ring-resonator frequency response [amplitude, ψ_r, vs. frequency]

(a) $\gamma = 0.95$; $Q = 2.3 \times 10^5$
(b) $\gamma = 0.85$; $Q = 7.3 \times 10^4$
(c) $\gamma = 0.6$; $Q = 2.3 \times 10^4$

radius = 0·5 mm
$n_e = 1.523$
$\lambda_0 = 632.8$ nm

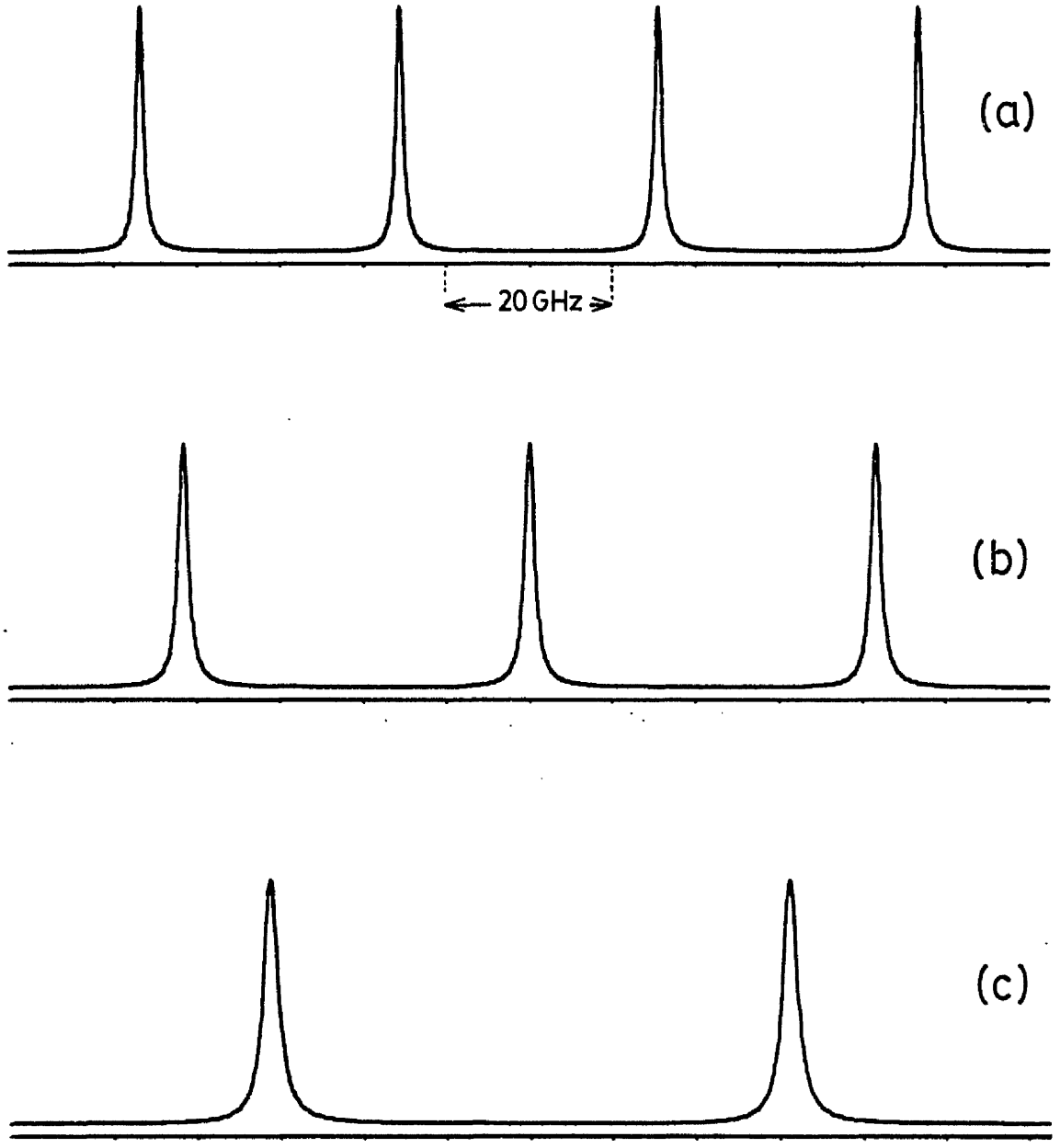


Fig. 3·5 Ring-resonator comb frequency response [as Fig. 3·4]

(a) R = 1·0 mm ; Q = $4{\cdot}6 \times 10^5$
(b) R = 0·75 mm ; Q = $3{\cdot}5 \times 10^5$
(c) R = 0·5 mm ; Q = $2{\cdot}3 \times 10^5$

$\gamma = 0{\cdot}95$

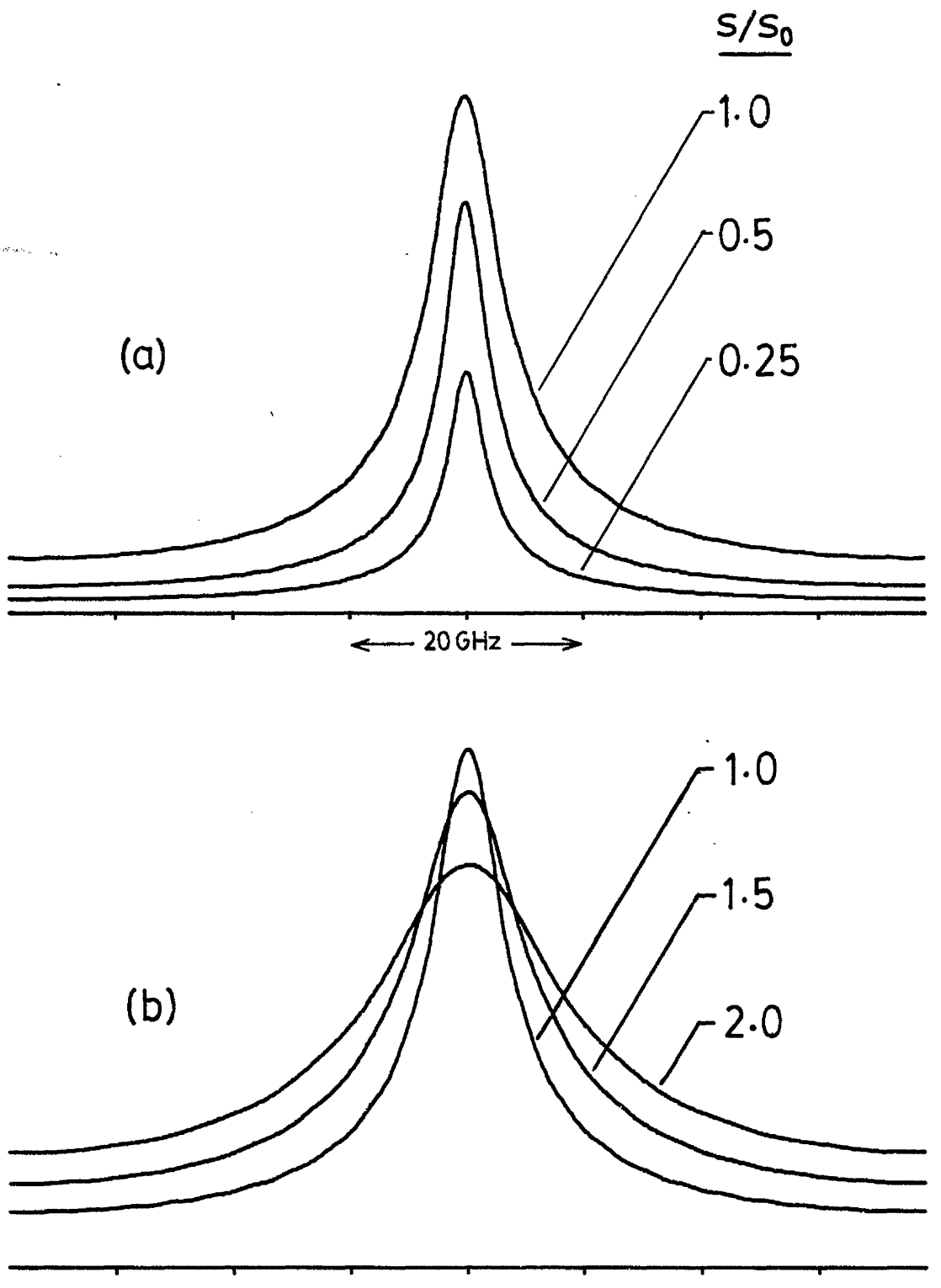


Fig. 3·6 – As for Fig. 3·4, but with input:
(a) under-coupled; (b) over-coupled
s is the actual, and s_0 the optimum, interaction length. γ is fixed at 0·9.

For example: a loss of 1 dB per circuit gives a value of γ of 0.891. This gives an acceptance ratio of 18.8 dB.

3.3 Losses in Optical Waveguides

It is clear from the previous section, that a major problem in the design of ring resonators concerns the input coupler. It is necessary to be able to produce a coupler with a pre-determined coupling coefficient. Of equal importance is the related problem of waveguide losses. Unless the ring transmission factor, γ, is known, the optimum coupling cannot be designed. Also, unless γ is close to unity (low loss), the device will be of very poor quality and the required input coupling impracticably large.

Straight Guide Losses

Losses exhibited by both straight and curved guides may be due to many distinct factors which include:

a) Scattering due to material bulk defects

b) Scattering due to imperfectly smooth boundaries

c) Absorption by the guide material.

Such imperfections generally cause coupling between the modes of a multimode waveguide and this contributes to the loss mechanism. A large mode-spacing helps to reduce such intermode coupling.

In general, these losses are not amenable to theoretical treatment and, in this work, will simply be measured.

Curvature Losses

The loss which occurs only at a change of curvature[23],[25] is not important for the circular ring resonator, therefore this section will be limited to a consideration of constant curvature losses.[23]-[33]

The propagation in a guide of constant curvature may be expressed:

$$\psi = \psi_0 e^{-j(\beta z)} . e^{-\alpha z}$$

$$= \psi_0 e^{-j(\beta - j\alpha)z}$$

where $\alpha = \alpha_L + \alpha_c$ is the attenuation constant, to which both linear and curvature losses contribute. $\beta - j\alpha$ may be considered as a complex propagation constant.

3.4 The Constant Curvature Loss

When a stripe waveguide takes the form of a curve in the x-z plane of constant radius, R_2 (the outer boundary radius), the tangential propagation properties λ_e and β (and hence n_e) become functions of the radial coordinate, ρ. For the guided wave to follow the curve it must have a constant angular phase-velocity, Ω_e, and a "wave-angle" rather than wave-length.

$$\lambda_e(\rho) = \frac{\lambda_{e_0} \rho}{R_2} \quad ; \quad n_e(\rho) = \frac{n_{e_0} R_2}{\rho} \qquad (3.4.1)$$

where λ_{e_0} and n_{e_0} are the values at the outer boundary of the waveguide.

At some radius (the critical radius $\rho = \rho_c$) $n_e(\rho)$ will equal the substrate, refractive index, n_4 [See Fig. 2.8 (a)].

$$n_e(\rho_c) = n_4 \qquad \text{where} \quad \rho_c > R_2$$

This is the usual cut-off criterion, as discussed in 2.1. For $\rho > \rho_c$ the evanescent field converts to a radiation field propagating tangentially to the arc $\rho = \rho_c$, and energy is lost.

From 3.4.1 $$\rho_c - R_2 = R_2\left[\frac{n_{e_0}}{n_4} - 1\right] \qquad (3.4.2)$$

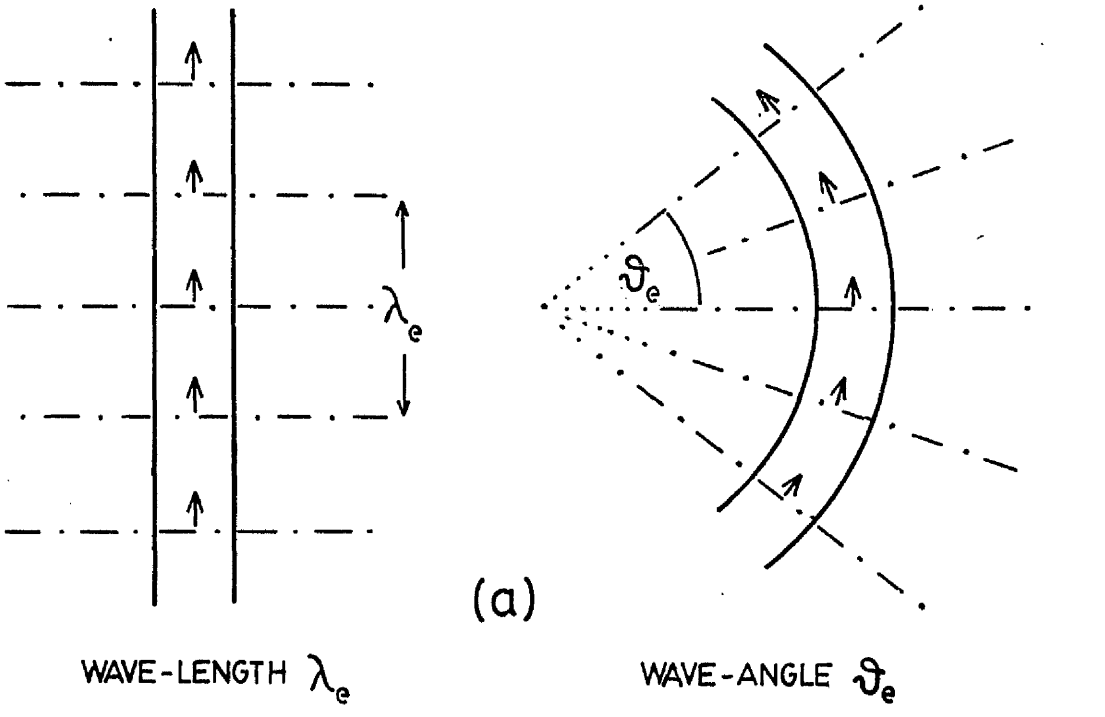
This loss is dependent upon the evanescent field amplitude at ρ_c. Minimising it requires that $[\rho_c - R_2]$ and ξ_4 (the evanescent field decay constant) both be large, thus requiring a large R_2 and a large ratio $\frac{n_{e_0}}{n_4}$. The bending loss is thus doubly dependent upon this index ratio, through both ξ_4 and $[\rho_c - R_2]$.

Since n_e is now a function of ρ, and:

$$\xi_4 \simeq k_0 \sqrt{n_e^2 - n_4^2}$$

it may be inferred that the field amplitude functions are changed by the bend. They become less tightly bound for $\rho > R_2$ and more tightly bound for $\rho < R_2$. This tends to increase the loss and means that the straight-guide fields cannot be used in any bending-loss calculation with complete accuracy.

The inner guide boundary (at $\rho = R_1$) may become redundant if, for some radius ρ_1 (where $R_1 < \rho_1 < R_2$), $n_e(\rho_1)$ exceeds n_3 (the inner substrate index). Thus the Pill-Box resonator[21] becomes a possibility.



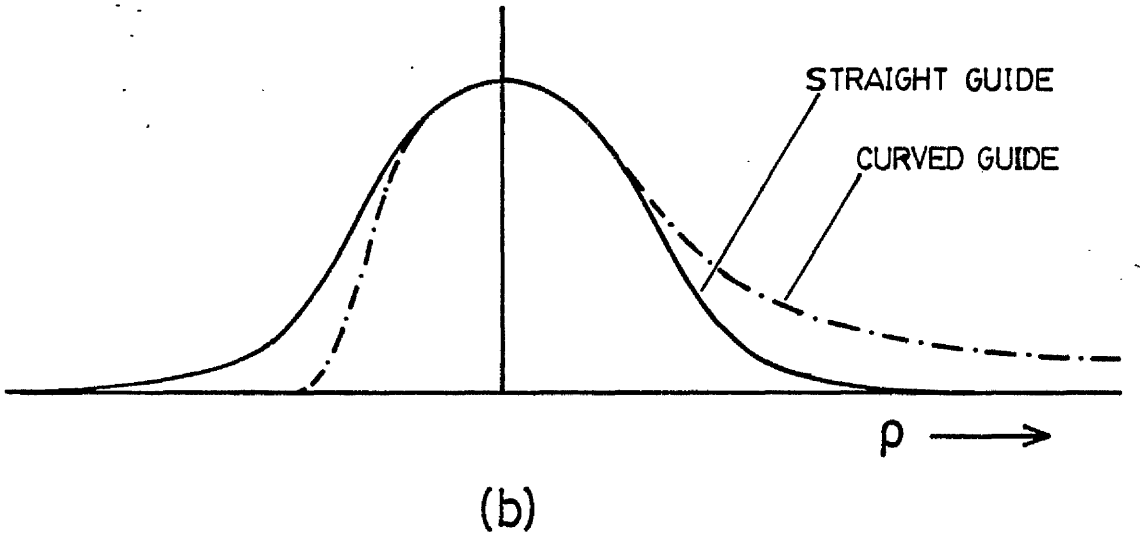


Fig. 3·7 The effect of guide curvature on :

(a) the phase fronts

(b) the field distribution

Analysis of the Curved Guide

Most published analytical methods concentrate on bent slab waveguides, or circular fibres (or other specific geometry).[24],[25],[26] Some use the straight guide fields to estimate the curvature attenuation constant (α_c). One of these - that of Marcatili and Miller[27] - as compared with the method of Marcatili[26] - is in error by a factor of 100 in the values of α_c. The problem is that the attenuation constant is very sensitive to the evanescent field spread. For this reason straight-guide approximations and bent slab/effective-index approaches are of dubious value.

The Index Transformation Method

The method to be outlined is based upon that of Heiblum and Harris[31],[32] in which a conformal mapping between complex planes is used to transform the curved waveguide into an equivalent straight structure. The method of (31) is valid only for bent slab waveguides and cannot be extended to stripe-waveguides unless the formalism of complex planes is abandoned. A direct transformation of variables, in three dimensions, is used here.

The equivalent, straight guiding structure has a complex propagation constant whose imaginary part gives α_c. Unfortunately there is no available method for calculating this. Nevertheless, the transformation alone gives valuable insight into the loss mechanism and will be presented in isolation.

Propagation in the curved stripe waveguide is described by the wave-equation in cylindrical coordinates (ρ, γ, θ):

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left[\rho\frac{\partial\psi}{\partial\rho}\right]+\frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\Theta^2}+\frac{\partial^2\psi}{\partial\gamma^2}+k_0^2n_1^2(\rho,\gamma)=0 \qquad (3.4.3)$$

where γ is the depth coordinate.

To find the equivalent straight guiding structure we require a change of variables which will transform the above into the wave-equation for a straight guide:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial z^2}+\frac{\partial^2\psi}{\partial y^2}+k_0^2n_2^2(x,y)=0 \qquad (3.4.4)$$

where the cartesian coordinates (x,y,z) are functions of (ρ,γ,Θ).

We try the transformation:

$$\begin{aligned}&(a)\quad z=R_2\Theta\\&(b)\quad x=R_2\ln\left[\frac{\rho}{R_2}\right]\\&(c)\quad y=\gamma\end{aligned} \qquad (3.4.5)$$

(a) conserves propagation distance at the waveguide;

(b) conserves the guide width as nearly as possible ($\delta x=\delta\rho$ for $\rho=R_2$), sets $x=0$ at the outer guide boundary and shifts the singularity at the centre of curvature to $x=-\infty$.

(a) and (b) follow Heiblum and Harris[31] and are known to map a circle to a straight line. (c) is simply a trial substitution.

Using 3.4.5, 3.4.3 becomes:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial z^2}+\frac{\rho^2}{R_2^2}\cdot\frac{\partial^2\psi}{\partial y^2}+k_0^2\frac{\rho}{R_2^2}n_1^2(x,y)=0$$

So the transformation must be modified. We try:

$$(c)\quad y=\frac{R_2}{\rho}\cdot\gamma \qquad (3.4.6)$$

In this case, the $\frac{\partial}{\partial \rho}$ term of 3.4.3 becomes very complex. However all the terms of this, except one, are of negligible magnitude, so 3.4.3 becomes, with very little error:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + k_0^2 \frac{\rho^2}{R_2^2} n_1^2(x,y) = 0 \qquad (3.4.7)$$

which is of the correct form (3.4.4).

Therefore, the transformed refractive index is:

$$n_2(x,y) = n_1(x,y) e^{x/R_2} \qquad (3.4.8)$$

The effect of this transformation upon the shape, dimensions and refractive index of the guide is illustrated in Fig. 3.8.

It may be noted that the transformed index profile has all the essential features of a prism-coupler (i.e. directional coupling into the unbounded region to the right). Clearly, guides which make good directional couplers will behave badly on bends - and vice-versa - since coupling and bend-loss are thus shown to be closely related phenomena. However, the waveguide bend would seem to be more critically dependent upon the index difference $n_e - n_4$ than the directional coupler. The coupling gap of the latter is fixed, whereas that of the former varies with this index difference.

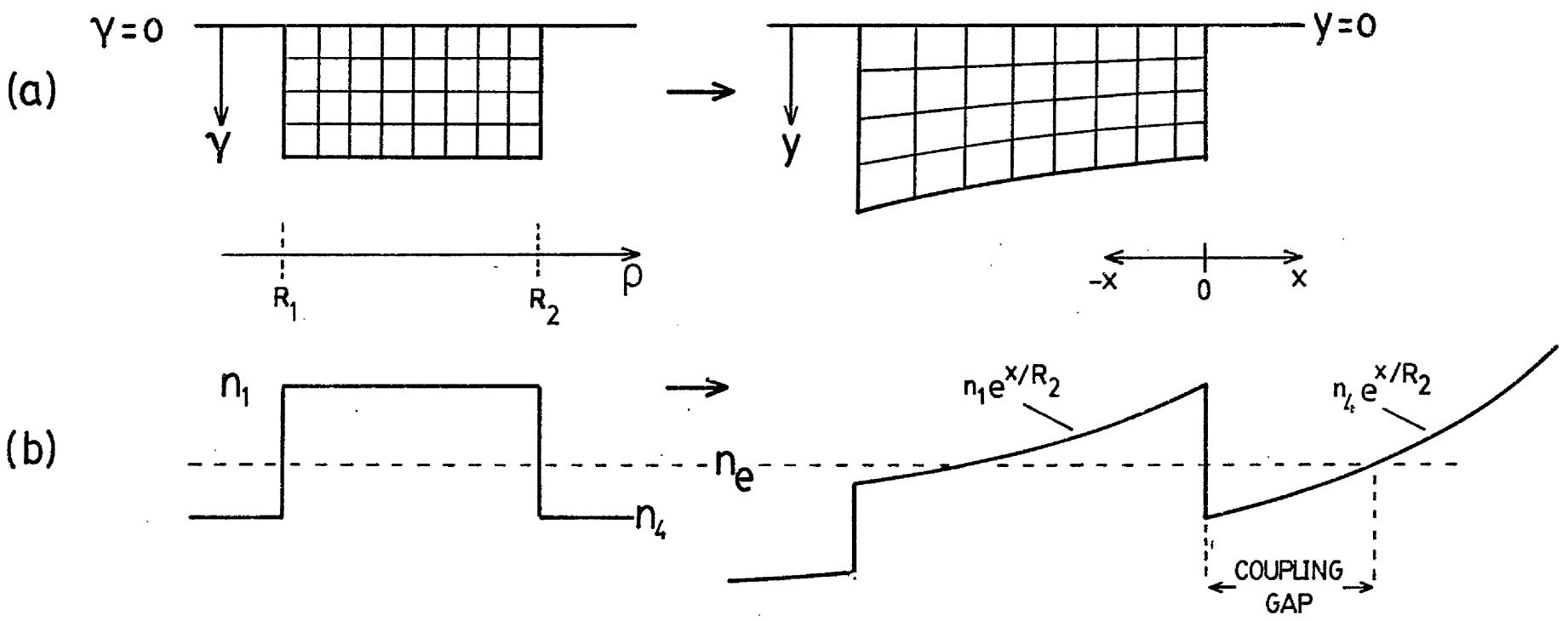


Fig. 3.8

The transformation of a curved guide to the equivalent straight guide, showing the effect on: (a) shape

(b) refractive index.

3.5 The Choice of Waveguide

In this chapter all the concepts fundamental to ring resonator design have been introduced. On this basis the waveguiding requirements may be summarised as follows:

1) Single mode guidance is highly desirable since each mode has its own coupling coefficient (K), bend loss and resonant frequencies. The presence of higher order modes would reduce the selectivity of the filter as well as reducing efficiency. To obtain single mode guidance it is advantageous to use narrow guides with a large index difference to give a large lateral mode separation.

2) A large directional coupling coefficient is required. If a ring transmission factor (γ) of 0.9 were achieved, optimum coupling would require almost 30% of the transfer length (L). Since rings of less than 1 mm radius are envisaged, the available interaction length is small. Thus, the guides should be narrow and the index difference $n_e - n_s$ (where n_s is the inter-guide index) should be small.

3) Low bending loss requires that the index difference $n_e - n_s$ be large.

Clearly, there is some conflict between (2) and (3) and a compromise must be sought. Hence:

4) Some versatility in the fabrication is desirable, in particular as regards refractive index differences.

3.6 The Silver Ion Exchanged Waveguide

The silver ion-exchanged waveguide is found to offer a suitable compromise to the above requirements.

These waveguides are made by immersing a substrate of soda-lime glass in a bath of molten silver nitrate ($AgNO_3$). Sodium ions diffuse out of the glass into the melt and are replaced by silver ions, thus increasing the refractive index near the surface. The mechanism of ion-exchange is considered in some detail in Chapter 4.

Stripe guides are made by coating the glass surface with a diffusion mask, in which slit-like apertures are opened. The resulting channel guides, being separated by the substrate material rather than by air, make better directional couplers than ridge or rib guides. Nevertheless, the difference between the surface and substrate refractive indices is large , compared with that of other types of diffused guide. This is typically about 0.09, which allows bends of fairly small radius to be made.

All diffused channel guides have the desired fabrication versatility since it is always possible to reduce the index difference of the finished waveguide by "baking" it; that is, by bringing it up to the diffusion temperature in the absence of a diffusant source, thus allowing the concentration gradients to relax.

In view of this possibility, the diffused guide with the highest initial refractive index is used - the silver ion-exchanged guide, made using pure silver nitrate.

Waveguides may also be made using potassium ion exchange, by using silver nitrate diluted with sodium nitrate[49] or by diffusion of titanium into lithium niobate ($LiNbO_3$). These all have a rather

low index difference; and problems can arise with the latter because of the optical anisotropy of $LiNbO_3$.

Diffused guides, naturally, have smoother edges than ridge guides[54] and may be expected to have lower straight-guide losses. Strip-loaded guides[59] were not considered, due to their very weak, lateral confinement.

The rest of this thesis is mainly concerned with various aspects of the silver ion-exchanged waveguide made using a pure silver nitrate melt.

The Rationale

A preliminary study of stripe ion-exchanged waveguides by Gallagher[55] revealed that the experimental width-dispersion curve (i.e. n_e vs. diffusion aperture width, for lateral modes) disagreed quite violently with that predicted by the Effective Index method.

It was felt to be essential that an accurate, theoretical model of these waveguides be available:

(a) so that some non-empirical design (of the directional coupler in particular) might be possible.

(b) To show up if, and how, the waveguide behaviour differs from the ideal so that improvements might be made.

Since the Effective Index method was not felt to be entirely trustworthy in this case, work was started on a model based on numerical techniques. The first step is to obtain the refractive index profile of the diffused stripe waveguide. This necessitates numerical solution (by digital computer) of the relevant diffusion differential equation. The diffusion theory pertaining to this solution is discussed in Chapter 4, and the solution itself in Chapter 5.

Having solved the diffusion equation, it is then necessary to solve the wave equation in the resulting 2-D region of arbitrary refractive index. This general numerical method is described in Section 5.2.

P A R T I I

CHAPTER 4

THE THEORY OF ION EXCHANGE

4.1 The Structure of Glass

Glass is not a single chemical substance. It is a mixture of fused, inorganic oxides, the principal component usually being silica (SiO_2); there is a very great number of possible oxide formulations which could be classified as glasses.

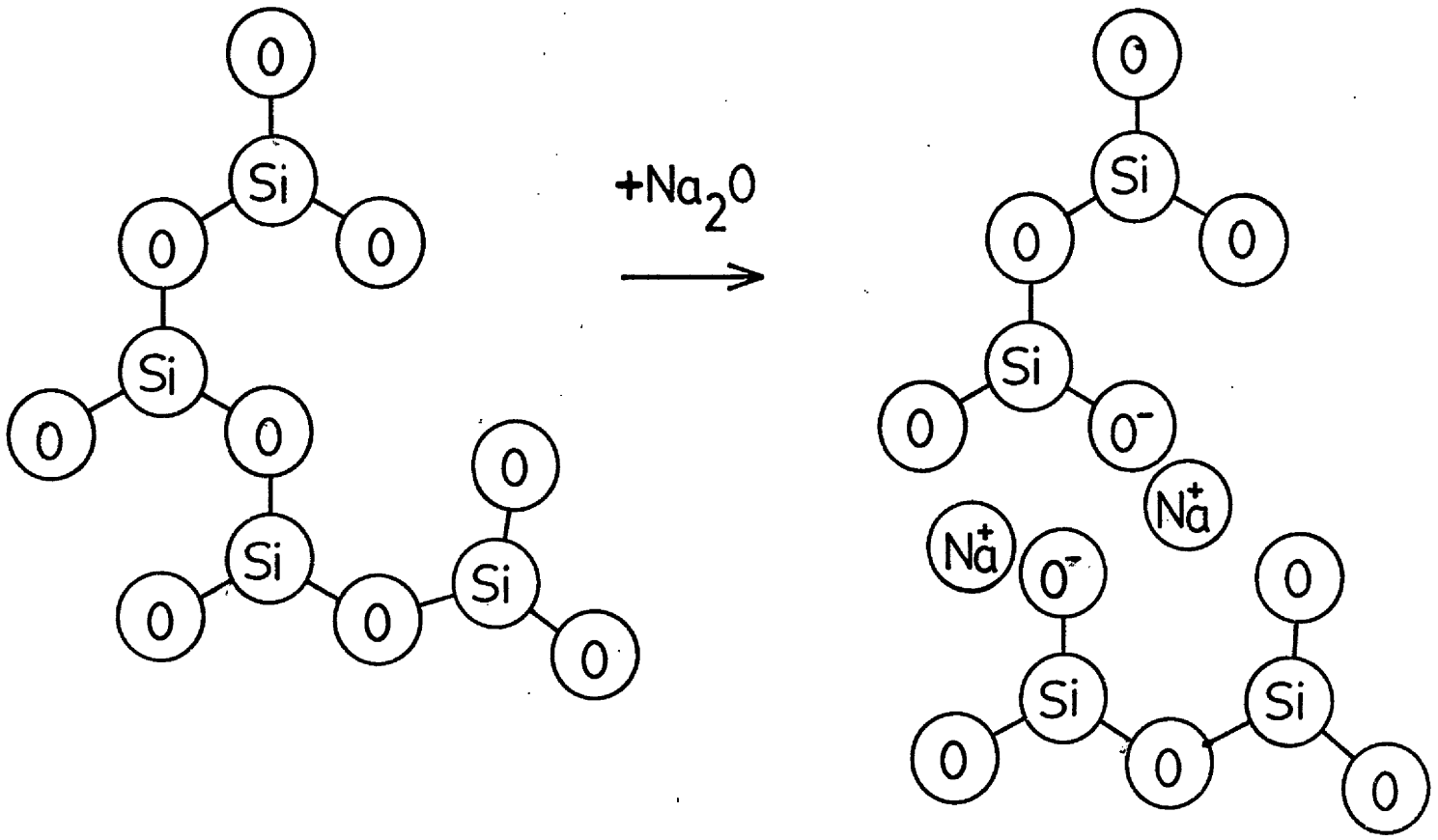
In this work interest is centred upon Soda-Lime Silica glass. This is used because of its ion-exchange properties and also because of its ready availability in the convenient and suitable form of microscope slides.

The structure of glass is not perfectly understood; it is believed to be that of a solid with the molecular disorder of a liquid frozen into its structure,[41] glass sometimes being described as a supercooled liquid.

The simplest type of glass is fused silica whose structure is illustrated in Fig. 4.1 (a).

The basic unit of silica is the SiO_4 tetrahedron which is held together by single covalent bonds. The bonding requirements of the oxygen atoms - that each be bound to two silicon atoms - cause the tetrahedra to link together into a random, three-dimensional SiO_2 network. Crystalline silica (quartz) is formed if the cooling is very slow and the minimum free-energy state is achieved before the structure is frozen.

Soda-Lime glass differs from fused silica in that it contains about 14% by weight of sodium oxide (Na_2O). The effect of this addition upon the resulting structure is suggested by Fig. 4.1 (b).



(a) Fused Silica

(b) Soda-Lime Glass

Fig. 4.1 Glass Structures

The extra oxygen allows the bonding requirements of many silicon atoms to be satisfied without recourse to cross-linking. Thus, the network is broken up and the single, bridging, oxygen atom of fused silica is replaced by two non-bridging oxygen ions, each carrying a single negative charge. The sodium ions (Na^+) reside in the vicinity of these fixed negative charge centres. This glass is less viscous and has a much lower melting point than fused silica.

An important feature of such glass is the relative freedom of the sodium ions. Given sufficient energy, they may escape from the potential wells in which they normally reside and move through the network. If the glass is immersed in a molten salt containing a different cation, sodium ions will leave the glass and the foreign cations will replace them.

Holland (41, p. 135), quoting earlier workers (Kraus and Darby - 1922) states that the cations from fused salts of Li, K, Cu, Pb, Zn, and Ag could all enter soda-lime glass to replace the sodium. However, with the exception of silver, the treated glass cracked on cooling. The silver ion has almost the same radius as that of sodium and so produces little strain in the glass.

The ion-exchanged waveguides studied in this work are produced by immersing a glass substrate in molten silver nitrate ($AgNO_3$). We therefore study the silver/sodium ion-exchange system with a view to computing, as accurately as possible, the refractive index profile of stripe ion-exchanged waveguides.

4.2 The Mathematics of Inter-Diffusion

The following assumptions and limitations are made.

1) We consider only the glass and ignore the melt. This is possible because of (2) below.

2) The glass surface is assumed to remain saturated with silver ions. This assumption is justified by Doremus[50] who found that the silver-ion partitions preferentially into the glass.

3) The self diffusion coefficients of silver and sodium in glass are assumed to be independent of concentration. This is also justified by the work of Doremus.[50]

4) The refractive index is assumed to vary linearly with silver concentration.

The process is considered in terms of two general monovalent ionic species labelled 'a' and 'b'. However the subscript 'a' may be taken to refer to silver and the subscript 'b' to sodium.

To ensure generality we consider glass of mixed composition, containing both ion species. The ionic concentrations (C_a and C_b ions/m^3) are functions of two space dimensions (x and y).

The glass is, initially, electrically neutral therefore the total positive ion concentration must equal the concentration of negative charge centres in the silica matrix. This is constant and is called C_k. Thus:

$$C_a(x,y) + C_b(x,y) = C_k \qquad (4.2.1)$$

Differentiating:

$$\text{grad } C_a(x,y) = - \text{grad } C_b(x,y) \qquad (4.2.2)$$

If the temperature is now raised to a level at which ion-exchange can begin, the concentrations become functions of time also. The

ionic fluxes ($\underline{\varphi}$) due to Fickian diffusion are:

$$\underline{\varphi}_a = - D_a \text{ grad } C_a$$
$$\underline{\varphi}_b = - D_b \text{ grad } C_b \qquad (4.2.3)$$

where D_a and D_b are the ionic self-diffusion constants. These are related to the ionic mobilities μ_a and μ_b by the Einstein relation:

$$D = \frac{kT}{e} \cdot \mu \qquad (4.2.4)$$

where k is Boltzmann's constant

T is the absolute temperature

e is the electronic charge.

Doremus[50] points out that 4.2.4 has been found, by several workers, to be inaccurate in glass, though his own findings are consistent with it. In (40) he states that phase separation in laboratory prepared glasses may be responsible for this. We (as did he) use commercial glass which contains some alumina to prevent phase separation. Thus 4.2.4 will be assumed to be accurate.

Since D_a is less than D_b, the heavy silver ion being the less mobile, it is clear that the two ionic diffusion fluxes cannot be equal. Since each ion carries a single positive charge, there will be a charge separation which is self-limiting; the resultant electric field equalising the total ionic fluxes.

The drift flux due to the electric field is (using 4.2.4):

$$\underline{\varphi}_e = DC \left[\frac{e}{kT}\right] \underline{E} \qquad (4.2.5)$$

where $\underline{E}$ is the total electric field strength. This includes any external fields which may be present.

Summing the two flux components for each ion-species gives the total flux.

$$\underline{\varphi}_a = - D_a\left[\text{grad } C_a - \left(\frac{e}{kT}\right)C_a\underline{E}\right]$$

$$\underline{\varphi}_b = - D_b\left[\text{grad } C_b - \left(\frac{e}{kT}\right)C_b\underline{E}\right] \qquad (4.2.6)$$

During ion-exchange diffusion, therefore, an electric field is always present. The overall neutrality of the diffusion system is retained but the glass will now contain regions of net positive and negative charge. The instantaneous volume charge-density, $\rho(x,y,t)$, is:

$$\rho = [C_a + C_b - C_k]e$$

So

$$C_a + C_b = C_k + \frac{\rho}{e} \qquad (4.2.7)$$

Comparison of 4.2.7 with 4.2.1 reveals that the latter is not strictly true in the dynamic state. However, examples on the final solution (see Appendix 2) confirm that $\frac{\rho}{e}$ is negligible by comparison with C_k, the difference being about nine orders of magnitude. Hence, for most purposes, 4.2.1 may be used with confidence as an approximation to 4.2.7.

Normalisation

It is convenient, at this stage, to normalise the concentrations with respect to C_k. We define:

$$u(x,y,t) = \frac{C_a(x,y,t)}{C_k}$$

$$w(x,y,t) = \frac{C_b(x,y,t)}{C_k}$$

Thus the exact relation 4.2.7 may be re-written:

$$u + w = 1 + \frac{\rho}{eC_k} \qquad (4.2.8)$$

We also define:

$$\alpha = 1 - \frac{D_a}{D_b}$$

α is a constant which describes the non-linearity of the final diffusion equation.

Hence, the normalised fluxes are (from 4.2.6):

$$\begin{aligned} \underline{\varphi}_u &= - D_a\left[\text{grad } u - \left(\frac{e}{kT}\right) u\underline{E}\right] \\ \underline{\varphi}_w &= - D_b\left[\text{grad } w - \left(\frac{e}{kT}\right) w\underline{E}\right] \end{aligned} \qquad (4.2.9)$$

In order to proceed further, we require the flux-continuity equations (from the divergence theorem):

$$\frac{\partial u}{\partial t} = - \text{div } \underline{\varphi}_u \qquad \text{and} \qquad \frac{\partial w}{\partial t} = - \text{div } \underline{\varphi}_w \qquad (4.2.10)$$

whence, from 4.2.8:

$$\frac{\partial}{\partial t}\left(\frac{\rho}{eC_k}\right) = - \text{div}(\underline{\varphi}_u + \underline{\varphi}_w) \qquad (4.2.11)$$

This gives the net current flow at the point in space concerned. Since $\frac{\rho}{eC_k}$ is very much less than unity we may say, with negligible error:

$$\text{div}(\underline{\varphi}_u + \underline{\varphi}_w) = 0 \qquad \text{and} \qquad w = 1 - u$$

Inserting the flux expressions (4.2.9), eliminating w and grad w and using the definition of α above, we obtain:

$$\text{div}[\alpha \text{ grad } u + \frac{e}{kT}\cdot(1 - \alpha u)\underline{E}] = 0 \qquad (4.2.12)$$

It must be noted that, because the divergence operator does not have a unique inverse, it may not be concluded, in general, that the operand of 4.2.12 is itself zero, although this is one possible solution. The conditions for which this solution may apply are considered in section 4.4.

In addition to 4.2.12 we have, from 4.2.9 and 4.2.10:

$$\frac{\partial u}{\partial t} = D_a[\nabla^2 u - \frac{e}{kT}(u \text{ div } \underline{E} + \underline{E}.\text{grad } u)] \qquad (4.2.13)$$

Similarly, expanding 4.2.12 gives:

$$\nabla^2 u = \frac{e}{kT}[E.\text{grad } u - \frac{1}{\alpha}(1 - \alpha u)\text{div } \underline{E}] \qquad (4.2.14)$$

Thus, eliminating $\nabla^2 u$ from 4.2.13, we find:

$$\frac{\partial u}{\partial t} = - \frac{D_a}{\alpha}\cdot\frac{e}{kT} \text{ div } \underline{E} \qquad (4.2.15)$$

but, from Maxwell's first equation:

$$\begin{aligned} \rho &= \text{div}(\varepsilon\underline{E}) \\ &\simeq \varepsilon \text{ div } \underline{E} \end{aligned} \qquad (4.2.16)$$

if the permittivity (ε) varies slowly.

Thus, the rate of change of silver ion concentration is seen to be proportional to the instantaneous charge density.

4.3 The Control System Model

In order to obtain a practical solution, we neglect the charge-density term when it occurs as the difference of two nearly equal quantities. However, in reality this term is a very important factor in the ion-exchange process. By tracing this physical process, we find that it behaves like a feedback control system and may be illustrated as such (see Fig. 4.2). The stages of the control loop are as follows:

1) The concentrations and their gradients constitute the system "input". To these are added an electric field term, fed back from the "output". This gives the fluxes of 4.2.9.

2) The fluxes are summed, converted to a scalar quantity by means of the divergence operator (4.2.10) and integrated with respect to time to give the charge density (4.2.11).

3) The loop branches at this point. $\underline{E}$ and $\frac{\partial u}{\partial t}$ both derive from ρ according to 4.2.16 and 4.2.15. Thus, a further integration gives u, and some suitable inverse-divergence operation gives $\underline{E}$. i.e.

$$\underline{E} = \frac{1}{\varepsilon} \operatorname{div}^{-1}(\rho)$$

Both terms are then fed back to the various input stages.

The Inverse Divergence Operation

As has already been noted, the divergence operator has no unique inverse. However, for the specific case of Maxwell's first equation, the electric field may be derived from the charge density function by summing the voltages, at a point (a,b), due to all the elemental charges dQ:

$$dQ = \rho(x,y)\,dx\,dy\,dz$$

The voltage at (a,b) due to dQ is dV

$$dV = \frac{dQ}{4\pi\varepsilon\sqrt{(x-a)^2 + (y-b)^2 + z^2}}$$

The total voltage at (a,b) is:

$$V(a,b) = \frac{1}{4\pi\varepsilon} \iiint_{-\infty}^{\infty} \frac{\rho(x,y)\,dx\,dy\,dz}{\sqrt{(x-a)^2 + (y-b)^2 + z^2}}$$

The electric field $\underline{E}$ = - grad V = $-\left(\frac{\partial V}{\partial a}, \frac{\partial V}{\partial b}\right)$

Hence, differentiating with respect to a and b and then performing the z-integral, we obtain:

$$\text{(i)} \quad E_x(a,b) = \frac{-1}{2\pi\varepsilon} \iint_{-\infty}^{\infty} \frac{\rho(x,y)\,.\,(x-a)}{(x-a)^2 + (y-b)^2}\,dx\,dy$$

$$\text{(ii)} \quad E_y(a,b) = \frac{-1}{2\pi\varepsilon} \iint_{-\infty}^{\infty} \frac{\rho(x,y)\,.\,(y-b)}{(x-a)^2 + (y-b)^2}\,dx\,dy \qquad (4.3.8)$$

In this model of the ion-exchange process every point in the x-y plane is governed by a system like that of Fig. 4.2. It is linked to adjacent points via the space derivatives, and to every other point by the above inverse divergence operation.

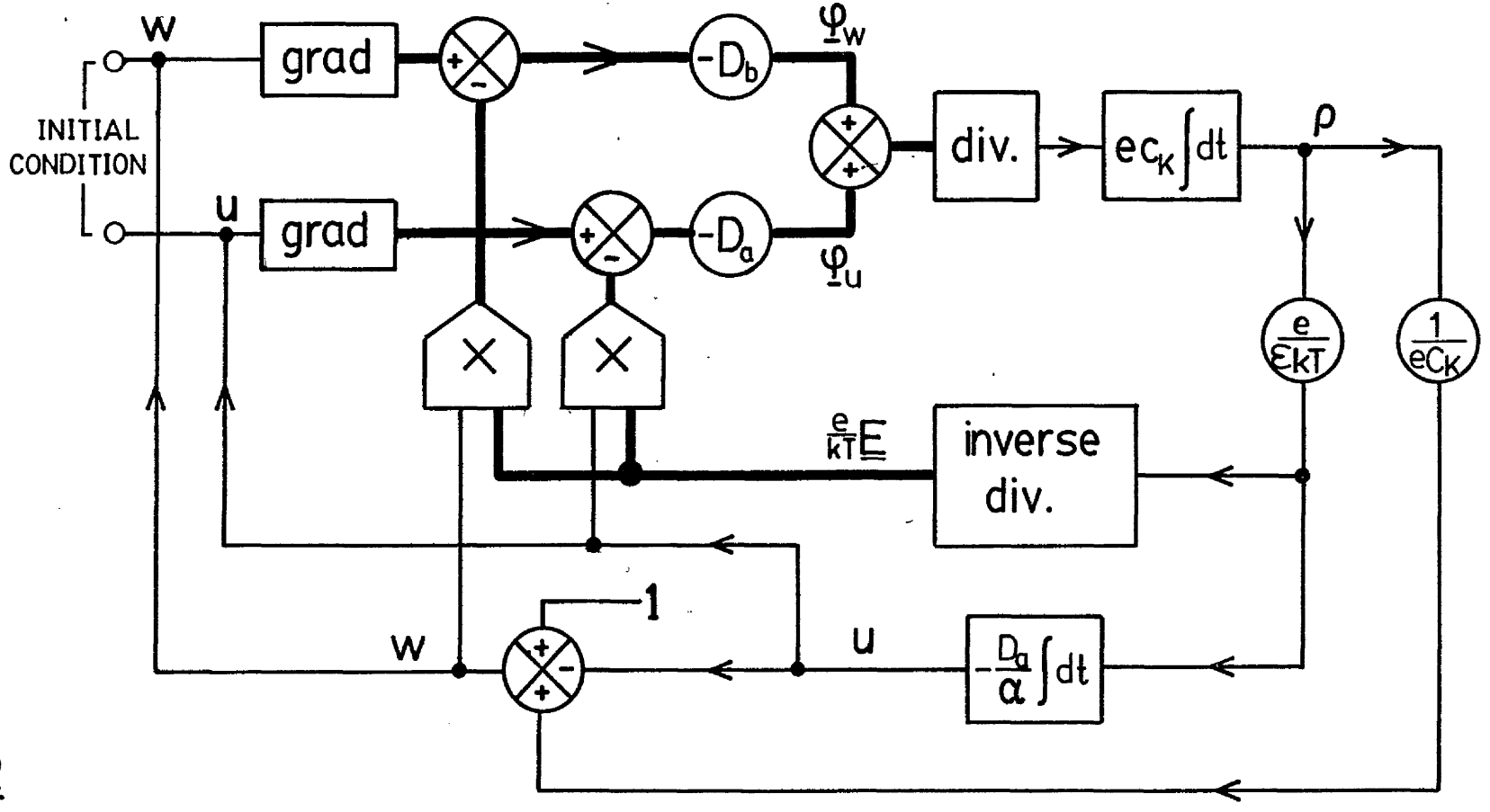


Fig 4·2

Control system model of ion-exchange

—— scalar

▬▬ vector

4.4 Solution in an Unbounded Medium

Before even a numerical solution of this diffusion problem can be attempted, it must be reduced to a partial differential equation in terms of u(x,y,t) alone; $\underline{E}$ and div $\underline{E}$ must be eliminated between equations 4.2.12 and 4.2.15. This is only possible if 4.2.12 can be "integrated" to reverse the divergence operation. As we have seen, this cannot be done directly and, in general, the operand of the divergence must be derived, in a more or less circuitous manner, using other criteria. As with the derivation of $\underline{E}$ from ρ (section 4.3) the basic physics of the situation must be considered.

4.2.12 is of form:

$$\text{div}(\underline{a} + \underline{b}) = 0 \qquad (4.4.1)$$

where vector $\underline{a} = \alpha \text{ grad } u$

and $\underline{b} = \frac{e}{kT}(1 - \alpha u)\underline{E}$

In an unbounded medium, with no externally imposed electric fields, grad u gives the direction of the Fickian flux components (4.2.3). The electric field, $\underline{E}$, which arises in response to the imbalance of these fluxes in order to restore the balance, will naturally be colinear with grad u. This balance being restored, $\underline{\varphi}_u$ will equal $-\underline{\varphi}_w$ which implies that $\underline{a} + \underline{b} = 0$.

Thus, the simplest possible solution of 4.4.1 applies in this case, giving:

$$\frac{e}{kT}\underline{E} = -\frac{\alpha}{1 - \alpha u} \text{ grad } u \qquad (4.4.2)$$

and from 4.2.15 with the above:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[\frac{D_a}{1 - \alpha u}\frac{\partial u}{\partial x}\right] + \frac{\partial}{\partial y}\left[\frac{D_a}{1 - \alpha u}\frac{\partial u}{\partial y}\right] \qquad (4.4.3)$$

The term $\frac{D_a}{1 - \alpha u}$ may be described as a concentration dependent inter-diffusion coefficient.

The above solution will not, in general, apply if externally imposed fields are present. In that case the direction of $\underline{E}$ will not be related to that of grad u; vectors $\underline{a}$ and $\underline{b}$ of 4.4.1 will not be colinear and $\underline{a} + \underline{b}$ can never be zero. A particular example of this (the conducting plane boundary) is considered in section 4.6.

4.5 The Insulating Diffusion Barrier

A practical solution of 4.4.3 requires information regarding boundary conditions, an important case being the diffusion barrier. Physically, this is an abrupt transition from the diffusion medium to a medium of zero diffusion coefficient which the diffusant cannot enter. An example of this is the diffusion mask which defines the width of a stripe waveguide; it is assumed to be electrically insulating and neutral in this section.

We define the diffusion mask (and the substrate surface) to be at the plane y = 0. The definitive equation for such a barrier is that the y-component of flux at the barrier be zero.

i.e. $$\varphi_{u_y} = 0 \quad \text{at} \quad y = 0 \qquad (4.5.1)$$

Thus, from 4.2.9

$$\frac{\partial u}{\partial y} = \frac{e}{kT} u E_y \quad \text{at} \quad y = 0 \qquad (4.5.2)$$

Since the barrier is electrically passive and insulating, the solution

of section 4.4 should be applicable. Substituting 4.4.2 in 4.5.2 we obtain:

$$\frac{\partial \bar{u}}{\partial y} = \frac{\alpha u}{1 - \alpha u} \frac{\partial u}{\partial y}$$

which is sensible only for the trivial solution:

$$\frac{\partial u}{\partial y} = 0 \quad \text{and} \quad E_y = 0 \quad \text{at} \quad y = 0 \qquad (4.5.3)$$

This is called the "reflecting boundary condition" because it is mathematically identical to the case of an infinite, uniform medium whose concentrations are symmetrical about the plane y = 0.

4.6 The Metallic Diffusion Barrier

The electrically conducting diffusion barrier is of some practical importance due to the use of aluminium as a diffusion mask. In this case, the electrostatic boundary condition causes severe disturbance of the electric field near the surface. This boundary condition is that all electric field components tangential to the conducting surface must vanish; charges move within the conductor, thus setting up a reaction field to ensure this.

The reaction field is given by the charge reflection property of a conducting plane, by which the conductor is replaced by the negative, mirror-image of the charge density function. The effect on the electric field at the boundary (y = 0) may be quantified using the "inverse divergence" operation of 4.3.

From 4.3.8 (with b = 0) we find that any function, $\rho(x,y)$ which is anti-symmetric about the plane y = 0 gives:

$$E_x(x,0) = 0 \quad \text{and} \quad E_y(x,0) = 2E_{y_0}(x,0)$$

where $E_{y_0}(x,y)$ is the value obtained for 4.3.8 (ii) if the y-integral is only semi-infinite.

Thus, although E_x is cancelled by the conducting plane, E_y is enhanced by it.

This unquestionably renders the solution of section 4.4 invalid. The direction of $\underline{E}$ is dictated by the conducting boundary orientation and is thus unrelated to that of grad u; $\underline{E}$ and grad u are not colinear so the differential equation cannot be simplified any further than 4.2.13 and 4.2.15.

Discussion of the Conducting Barrier Problem

An exact solution of the diffusion problem with a conducting boundary is impractical; however it seems likely that, despite its profound effect upon the electric field, the conducting boundary may have little effect upon the final diffusion profile.

Of itself, a conducting plane has no effect upon a given charge density, ρ; therefore $\underline{E}$ must be changed in such a way that div $\underline{E}$ is not changed. In terms of the control-system model of the ion-exchange process, the "inverse divergence" operation is modified by a conducting boundary.

But we see. from 4.2.15, that:

$$\frac{\partial u}{\partial t} \propto \text{div } \underline{E} \propto \rho$$

We consider two stripe waveguide diffusions, identical save that one uses a conducting, and the other an insulating diffusion mask. If both have identical charge density functions, the above considerations suggest that both diffusions will proceed at the same rate. However, it is not clear that both diffusions would develop identically from the same initial conditions.

We could, for example, conduct a "thought experiment" in which an insulating boundary is instantaneously made conducting during diffusion. 4.2.14, re-written, states:

$$\nabla^2 u - \frac{e}{kT}[\underline{E}.\text{grad } u] = -\frac{1}{\alpha}(1 - \alpha u)\text{div } \underline{E} \qquad (4.6.1)$$

Before the transformation

$$\underline{E}.\text{grad } u = E_x \frac{\partial u}{\partial x} \quad \text{at the boundary.}$$

After the transformation, $\underline{E}$ changes immediately, but div $\underline{E}$ and all the concentrations respond (if at all) much more slowly. Thus the only immediate change is:

$$\underline{E}.\text{grad } u \longrightarrow 0 \quad \text{at the boundary.}$$

and the equation no longer balances.

The equations do not take account of such transient conditions and the fluxes would be unbalanced until such concentration changes as would balance 4.6.1 occur. The implication is that $\frac{\partial u}{\partial y}$ is non-zero at a conducting barrier, the "reflecting" concentration distribution being incompatible with it.

Nevertheless, it seems likely that such differences as exist between the two diffusion profiles will be small and will be confined to the vicinity of the diffusion mask. The reasons are as follows:

1) The diffusion mechanism is not disabled by the conducting boundary. Although the x-direction flux components remain unbalanced (because E_x is reduced), the y-direction flux components possess the opposite imbalance (because E_y is enhanced). Thus the charge density is limited and silver ions may flow, unimpeded, into the vicinity of the mask, albeit by an indirect route.

2) Elsewhere, in each system (i.e. far from the mask) the diffusion rates will be identical. If (for example) $\frac{\partial u}{\partial t}$ were smaller near the conducting mask than near the insulating mask, the resultant increase in the concentration gradient would largely correct this (by means of the outer feedback loop of Fig. 4.2).

3) Unlike a true external field, the conducting boundary neither adds nor subtracts anything (either ions or motive force) to or from the diffusion system. Thus, there is unlikely to be any difference in the quantity of silver in each waveguide, only a small difference in its distribution.

Lacking an exact solution for this system, therefore, we use the reflecting boundary condition to model both insulating and metallic diffusion masks.

4.7 The u = 0 Boundary Condition

The boundary condition: $u = 0$ at $y = 0$ indicates the presence of an unlimited source of sodium ions and a sink for silver ions. It can be used to define the width of a stripe waveguide in the mathematical model of Chapter 5. Since silver is actually removed by this boundary, it may be taken to represent a boundary perturbation, far in excess of that introduced by a metallic diffusion mask. The feedback connection mentioned in (2) above is, to an extent, disabled by the extraction of silver ions ($\frac{\partial u}{\partial t}$ is forced to zero). Nevertheless, as is shown in Chapter 5, the effect of this extreme perturbation is smaller than might be expected, being a reduction in the effective guide width of about 0.2 μm in a 3 μm wide waveguide.

4.8 The Full Ion-Exchange System

It must be emphasised that all the above boundary considerations take account only of direct effects upon the inter-diffusion mechanism. Other indirect effects caused by interaction with the molten silver nitrate, for example, are not considered here.

A full model of the ion-exchange system must take account of the melt, though if the glass surface remains saturated at all times this surface layer may be considered as the silver ion source and only half of the system solved. For these conditions the half-system solution is exact, and $\underline{E}$ may be found using 4.4.2.

If the problem is one-dimensional (a slab waveguide), the charge density is:

$$\rho = \frac{\partial}{\partial y}(\epsilon \underline{E})$$

$\underline{E}$ is found to decrease monotonically with depth (see Appendix 2) and it follows that, during ion exchange, the glass possesses a net negative charge, also indicating the presence of a positive space charge in the melt, near the glass surface.

Thus, the glass may be regarded as an electrode which, when immersed in an electrolyte (the molten $AgNO_3$) develops an electrode potential. In this it is analogous to a corroding metal or, more nearly, to an ionic solution in contact (via a salt-solution bridge) with a different solution. The consequences of this will be considered in a later chapter.

CHAPTER 5

THE MATHEMATICAL MODEL

5.1 Solution of the Diffusion Equation

The computer program which is used to calculate the refractive index profile of a general ion exchanged waveguide is called DIF. It solves the non-linear partial differential equation:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[\frac{D_a}{1 - \alpha u}\frac{\partial u}{\partial x}\right] + \frac{\partial}{\partial y}\left[\frac{D_a}{1 - \alpha u}\frac{\partial u}{\partial y}\right] \qquad (5.1.1)$$

Details of the method and flow-charts of DIF may be found in Appendix 3.

In the finite-difference technique used[35)-(39)] each derivative of 5.1.1 is replaced by its finite-difference approximation.

e.g. $\frac{\partial u}{\partial x}$ is replaced by $\frac{U(x+\delta x,y) - U(x-\delta x,y)}{2\delta x}$

The rectangular array U(x,y) represents a rectangular region of the x-y plane and contains point-values of normalised silver concentration (u) taken at intervals of δx and δy. In computational terms the values are given by U(i,j) (i and j are integers) where:

$$x = (i - 1)\delta x \quad \text{and} \quad y = (j - 1)\delta y$$

Three such arrays are defined, to contain concentration values at three time levels: LU(i,j), U(i,j) and NU(i,j) represent times $t - \delta t$, t and $t + \delta t$ respectively. Thus:

$$\frac{\partial u}{\partial t} \longrightarrow \frac{NU(i,j) - LU(i,j)}{2\delta t}$$

When all the terms of 5.1.1 are thus replaced, the differential equation may be rearranged to give NU(i,j) in terms of LU(i,j), and U(i,j) and its four surrounding points (see A3.1.3). Assuming that LU and U are available, the whole of NU may be computed, point by point. LU is then discarded and replaced by U, U is replaced by NU, the time counter is incremented and the process repeated.

Since only one time level is initially available (given by the t = 0 boundary conditions) a two-level algorithm is used to compute the first few time-steps. Unlike the three-level method summarised above, the two level method is not unconditionally stable (instability is caused by error build-up) but requires a very small time step, δt. The normalised time step r is defined as:

$$r = \frac{D_a \, \delta t}{\delta x \, \delta y} \qquad (5.1.2)$$

Only if $\delta x = \delta y$ may these terms be so grouped.

For stability; $r < (1 - \alpha)/4$ from (A3.2.2)

However the three-level method can use a much larger time step, and this is necessary for computational efficiency. However, near the time origin, the very rapid changes require a small value of r, for the sake of accuracy, even with the three-level algorithm. For this reason, the program begins with $r = r_0/64$ (where r_0 is the desired final value) and r is doubled every 8 time steps until $r = r_0$. This gives a good compromise between speed and accuracy.

When the desired number of iterations is complete, NU is written into a data-file for use by other programs - the contour plotter, or the wave equation solution to be discussed in section 5.2.

Boundary Conditions

A fourth rectangular array, NB(i,j), contains a numerical code which dictates the boundary conditions. The boundary condition code is as follows:

0 No boundary.

1 $u = 0$; A source of sodium ions.

2 $u = u_0$; If $u_0 = 1$ this represents glass which is saturated with silver ions.

3 $\frac{\partial u}{\partial x} = 0$; Reflection about a y-direction boundary line.

4 $\frac{\partial u}{\partial y} = 0$; Reflection about an x-direction boundary line.

5 $\frac{\partial u}{\partial y} = \frac{\partial u}{\partial x} = 0$; Defined only where a 3 boundary meets a 4 boundary at a corner.

It is possible to define a boundary along any x or y-directed line, however only the four edges are normally used, the bulk of NB being filled with zeros. To save computer time, only half of a symmetrical structure is computed.

It is possible to start diffusion from some initial profile, which is input from a data-file. Thus, after some initial diffusion, the source boundary conditions may be changed, the output-file copied into the input-file and diffusion recommenced. Various 2 or 3 stage diffusion processes (such as the sourceless baking of a waveguide mentioned in section 3.6) may be simulated in this way.

Finally, the refractive index variation is computed, assuming the linear relation:

$$n(x,y) = n_s + (n_1 - n_s)u(x,y)$$

Acknowledgment

Thanks are due to Dr. C.D.W. Wilkinson who developed the first working version of DIF on which the refined version, described herein, is based.

5.2 Solution of the Wave Equation

The wave equation to be solved, in some region of arbitrary refractive index distribution, is the Helmholtz scalar wave equation (see Appendix 1):

$$\nabla^2_{xy} E_x + k^2 E_x = \beta^2 E_x \qquad (5.2.1)$$

Only Quasi-T.E. ($\underline{E} \simeq E_x$) modes are considered.

The computer program which solves 5.2.1 is called WAVE. It uses a variational technique which is fully described in Appendix 4. A brief summary of the method follows.

The variational expression (5.2.2) is a re-statement of the wave equation (5.2.1) which incorporates the fact that 5.2.1 must be true over all space.

$$F(E_x) = \beta^2 = \frac{\int_{-\infty}^{\infty}\int (\nabla^2 E_x + k^2 E_x).E_x \, dx\, dy}{\int_{-\infty}^{\infty}\int E_x.E_x \, dx\, dy} \qquad (5.2.2)$$

where k and E_x are functions of (x,y).

We consider the functional, $F(\psi)$, where ψ is a trial function (replacing E_x in 5.2.2) which may be varied. It can be shown that, if $\psi = E_x$, then the value of $F(\psi)$ is stationary and equals β^2. However, if

a stationary value is found, as ψ is varied

$$\text{i.e.} \qquad \frac{\partial F}{\partial \psi} = 0 \qquad (5.2.3)$$

it may only mean that the closest possible approach to $E_x(x,y)$ and β, for the form of function used, has been found. Therefore, we use a form of trial function which is capable of approximating to E_x as closely as required.

The Rayleigh-Ritz Method

The trial function, $\psi(x,y)$, is a series of orthogonal basis functions in two dimensions; each term of the series includes a coefficient, a_{ij}, which is varied to satisfy 5.2.3. The more terms the series has, the more closely may ψ approximate to E_x.

To minimise the number of terms necessary, we choose a set of orthogonal basis functions which closely resemble the modes of a waveguide. These are the parabolic cylinder (or Hermitian-Gaussian) functions which describe the modes of a laser resonator.[3] The parabolic cylinder function of order n is defined thus:

$$D_n(z) = (-1)^n e^{z^2/4} \cdot \frac{d^n}{dz^n} [e^{-z^2/2}]$$

Each term of the series is the product of an x-direction and a y-direction parabolic cylinder function. These must be scaled (via parameters ξ and η) and normalised (the norm of the function $D_i\left(\frac{x}{\xi}\right)$ is $\sqrt{2\pi}\, i!\xi$). The guide is assumed to be symmetrical about the y-axis, but not about the x-axis, therefore a zero-shift parameter, y_0, is required.

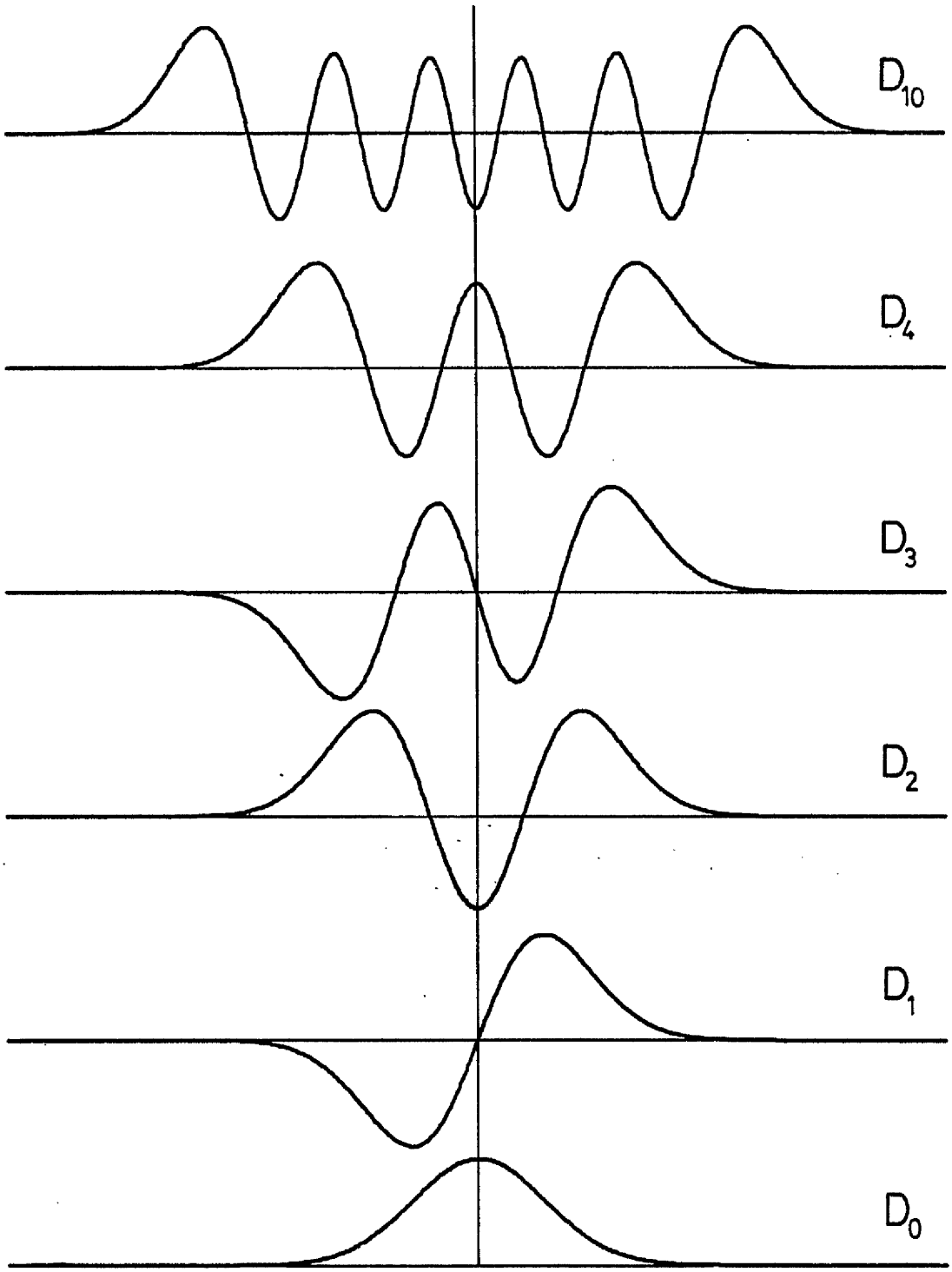


Fig. 5·1 Parabolic Cylinder Functions $D_n(x)$

Finally, the series, $\psi(x,y)$, is:

$$\psi(x,y) = \sum_{i=0}^{n}\sum_{j=0}^{m} a_{ij}(2\pi\xi\eta i!j!)^{-\frac{1}{2}} D_i\left(\frac{x}{\xi}\right) D_j\left(\frac{y - y_0}{\eta}\right) \qquad (5.2.4)$$

It may be noted that this is not the same as the product of two 1-D series which would have only n + m independent coefficients. The above, by contrast, has n x m independent coefficients; thus, the method is not limited to a rectangular geometry (see Section 2.3).

This series is substituted for E_x in 5.2.2 and $\frac{\partial F}{\partial a_{ij}}$ is set to zero, resulting finally in the matrix eigenvalue equation:

$$[\underset{\sim}{N} - \beta^2 \underset{\sim}{I}]\underset{\sim}{a} = 0 \qquad (5.2.5)$$

where $\underset{\sim}{N}$ is a symmetric square matrix (nm x nm), with elements N_{ijkl} derived from the numerator of 5.2.2 and given by A4.3.1. Since each series (5.2.4) in the product ψ x ψ requires independent counters, the extra counters k and l appear.

The matrix $\underset{\sim}{N}$ has nm possible eigenvalues (β^2) each with its corresponding eigenvector, $\underset{\sim}{a}$, which is a column matrix containing the values of a_{ij}.

Practical Details

The main task of WAVE is to assemble the matrix $\underset{\sim}{N}$, after which the eigenvalues and eigenvectors are computed by means of standard matrix manipulation routines. Since computation of each element of $\underset{\sim}{N}$ involves a surface integral, the program will be very slow unless the numbers of basis functions used, n and m, are minimised. $\underset{\sim}{N}$ is symmetric, so only

half of it need be computed. Also the waveguide is symmetrical so its modes will be either symmetric or antisymmetric, allowing exclusively even or odd order x-direction basis functions to be used. Thus, n is halved, without loss of accuracy. The program must be run twice to obtain all the modes, but the computer time is nevertheless cut by half.

The parameters η, ξ and y_0 are optimised by adjusting these until the lowest order matrix element, N_{0000}, is at a maximum. That is, the field is approximated by a single basis function in each direction ($\underset{\sim}{N} = N_{0000} = \beta^2$). The resulting parameter set is close to the optimum for rapid convergence of the series 5.2.3 and gives the best-fitting Gaussian function to the fundamental-mode field. Note that, with an infinite series (a complete orthogonal set), the field could be matched for any values of η, ξ and y_0. The above procedure simply chooses a set which gives the most rapid convergence, thus allowing a small number of terms to be used with acceptable accuracy.

A 1-D version of WAVE (called SLAB) is also available to deal with slab waveguides by the same method.

Acknowledgements

A technique, essentially the same as that described above, has been used by Matsuhara,[52] and Taylor[51] for solving various types of dielectric waveguide.

The program used in this work was first developed by Drs. C.D.W. and J.A.H. Wilkinson though some of the described features (e.g. the 1-D version) are due to the present writer.

5.3 Behaviour of the Diffusion Equation Solution

a) The Diffused Slab

The computer program DIF will compute the diffusion profile of an ion-exchanged slab waveguide if both vertical (y-direction) boundaries use the reflecting boundary condition $(\frac{\partial u}{\partial x} = 0)$ and the surface is a uniform silver ion source; the array containing the values of u need have only three columns.

A set of typical diffusion profiles is plotted (u vs. y) in Fig. 5.2 for six different values of the parameter α. These curves agree with those given by Crank [(35) p. 175] for a 1-D diffusion, solving 4.4.3 analytically. The curve for $\alpha = 0$ is the solution of a simple linear diffusion equation and follows the error-function complement.

It may be expected from the form of the inter-diffusion coefficient of 4.4.3, that as α tends towards unity the diffusion profile will tend towards an infinitely deep step-function. This is borne out by Fig. 5.2, and also by the dispersion curves of Fig. 5.3, (calculated using the W.K.B. method) which increasingly resemble those of Fig. 2.3 (for a step index slab) as α is increased.

b) Depth Scaling

In the dispersion curves of Fig. 5.3 the depth 'd' is taken to be that of the u = 0.5 concentration level. Thus, these curves do not reflect the increasing depth as α increases which is clearly seen in Fig. 5.2, but only the changing shape. This is equivalent to scaling the depth of each curve of Fig. 5.2 so that all intersect at u = 0.5. We consider this scaling.

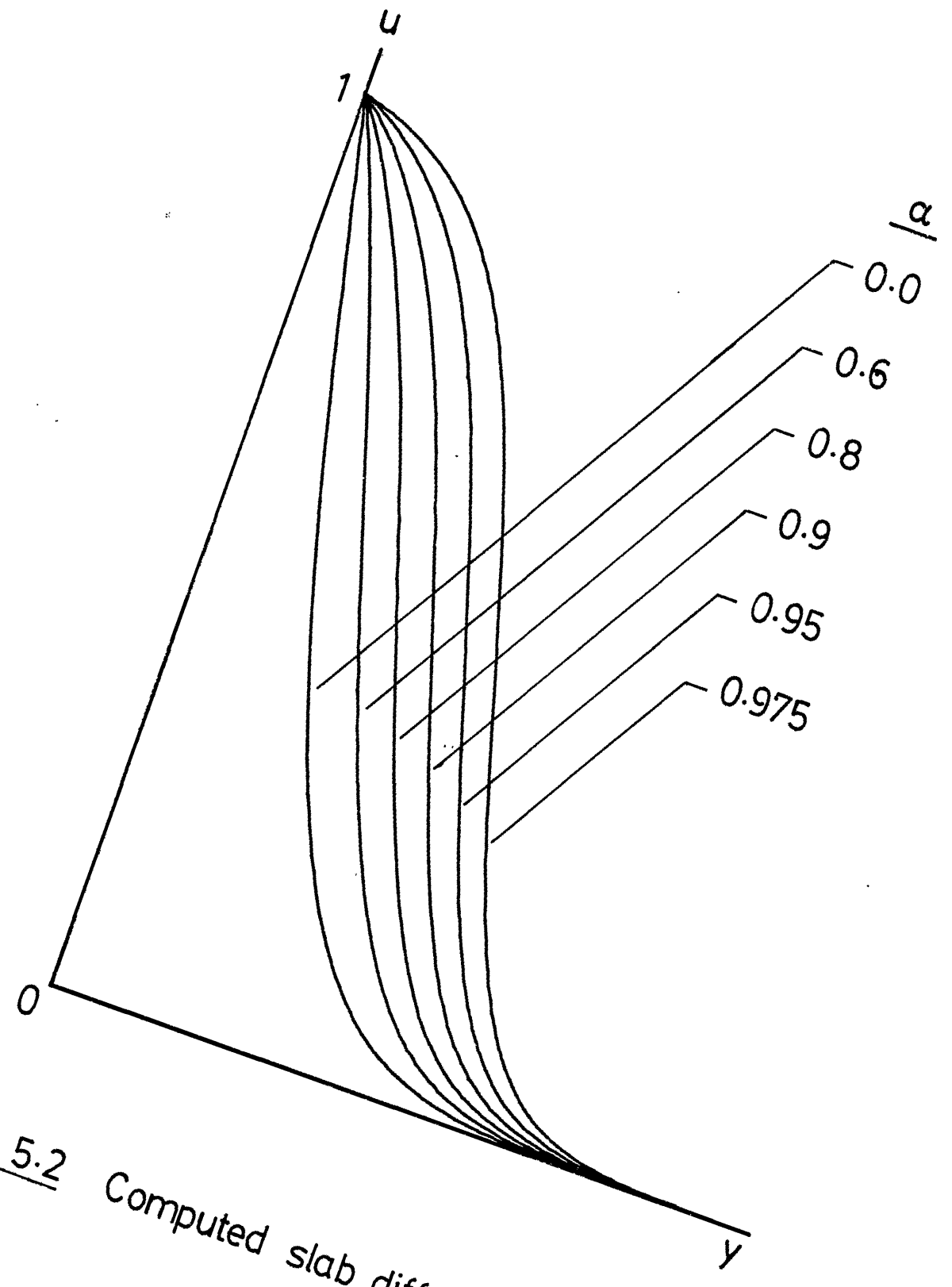


Fig. 5.2 Computed slab diffusion profiles

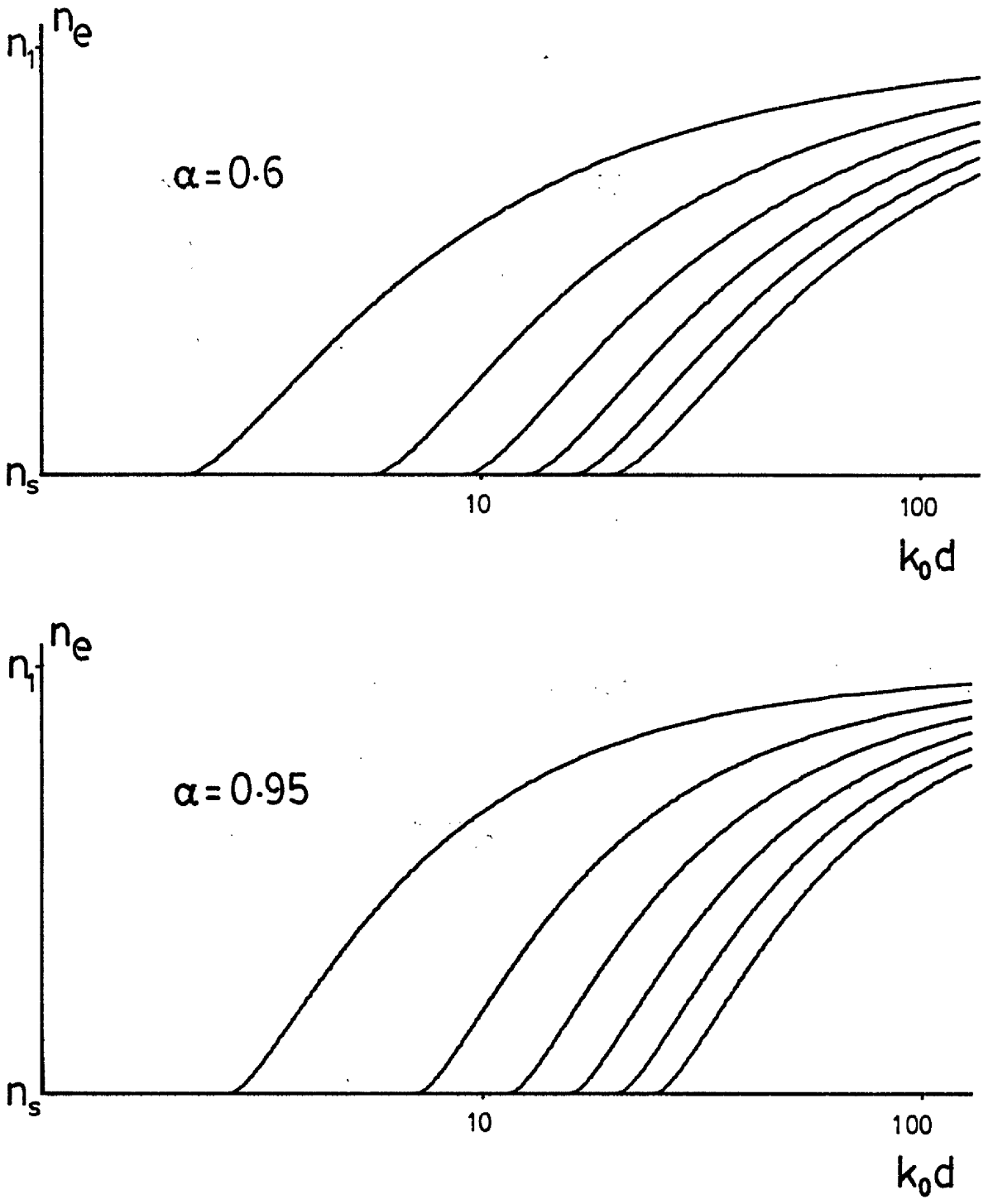


Fig. 5.3 Theoretical dispersion curves for ion-exchanged slab waveguides.
[T.E.]

The normalised diffusion time t_n is:

$$t_n = r_0 N_t$$

where N_t is the equivalent number of large time steps (r_0) used in the computation; r_0 is the normalised time step:

$$r_0 = \frac{D_a \delta t}{\delta y^2}$$

Thus, $$\frac{t}{\delta y^2} = \frac{t_n}{D_a}$$ where t ($= N_t \delta t$) is the real diffusion time.

If D_a is known and t_n is fixed by the computation, then the real diffusion time represented by the computed profile may be varied by changing the depth scaling term, δy; the shape of the profile is unaffected by t_n. Alternatively, if the real diffusion time is known - as with a real waveguide - then D_a may be found by using any known value of t_n and adjusting δy until the computed n_e values match the measured values.

In general the depth 'd' of an ion-exchanged guide varies as the square root of diffusion time:

$$d \propto \sqrt{D_a t}$$

c) <u>The Value of r</u>

According to the above, a constant value of t_n with varying r_0 and N_t should always give the same results. However, in practice a change of r_0 is found to change the results slightly - especially if α is high ($\alpha > 0.9$).

For example, a stripe waveguide calculation, with $\alpha = 0.97$, was performed for $r_0 = 0.075$. The same calculation, with $r_0 = 0.3$ gave a

somewhat different n_e value. With $r_0 = 0.5$, the process was still stable, though the profile seemed rather distorted, but with $r_0 = 0.7$ the profile was badly distorted.

When α is as low as 0.7, values of r_0 which exceed unity may be used with safety; however it was usual in this work to use $r_0 = 0.6$ with $\alpha = 0.7$. It is good practice to keep r_0 constant between computations whose results are to be compared.

d) Stripe Waveguide Diffusion Profiles

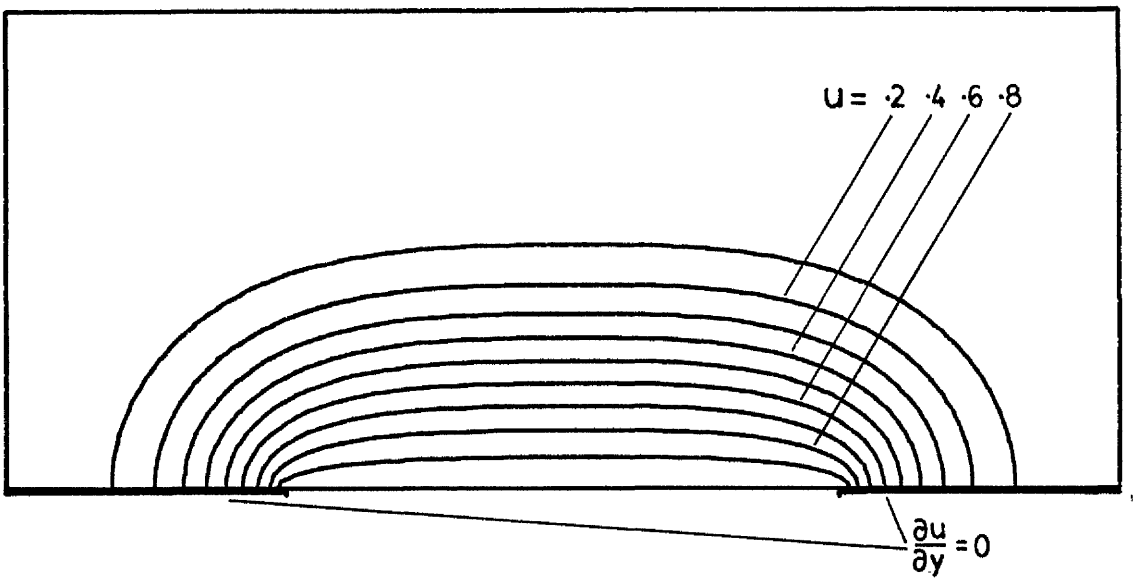
Fig. 5.4 shows the diffusion profiles of stripe, ion exchanged waveguides, plotted as contour maps of the x-y plane. The contours $u = 0.1, 0.2, \ldots 0.9$ are plotted, the surface being fixed at $u = 1.0$.

The computation of Fig. 5.4 (a) uses the reflecting boundary condition, $\frac{\partial u}{\partial y} = 0$ (see section 4.5), which is considered to represent accurately the electrically insulating diffusion mask. The computation of Fig. 5.4 (b) uses the boundary condition, $u = 0$ at $y = 0$, (see section 4.7) to represent the diffusion mask. This easily implemented boundary represents a sink for silver ions and is used to check the effect of a severe boundary perturbation.

In both cases, the array containing the values of u has 101 columns and 45 rows. The diffusion aperture width is $50\delta x$; $\alpha = 0.7$ and $t_n = 60$.

A real waveguide with a diffusion aperture width of 2 microns, diffused for 20 minutes at 215°C would have the proportions of 5.4 (a).

(a)



(b)

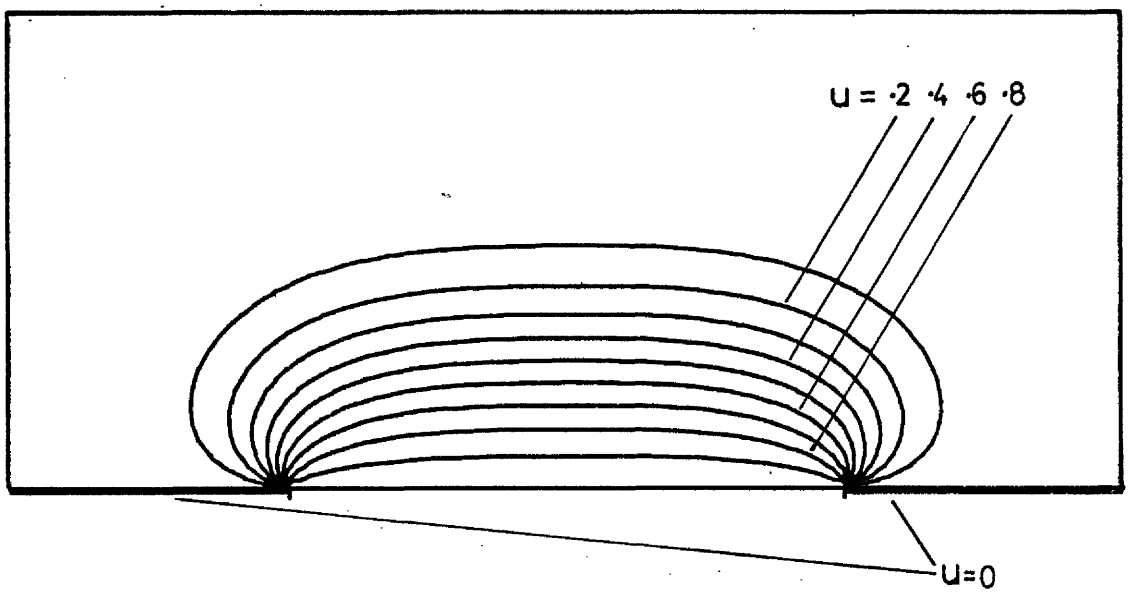


Fig. 5·4 Computed stripe diffusion profiles using: (a) the 'reflecting', (b) the 'u=0' diffusion masks

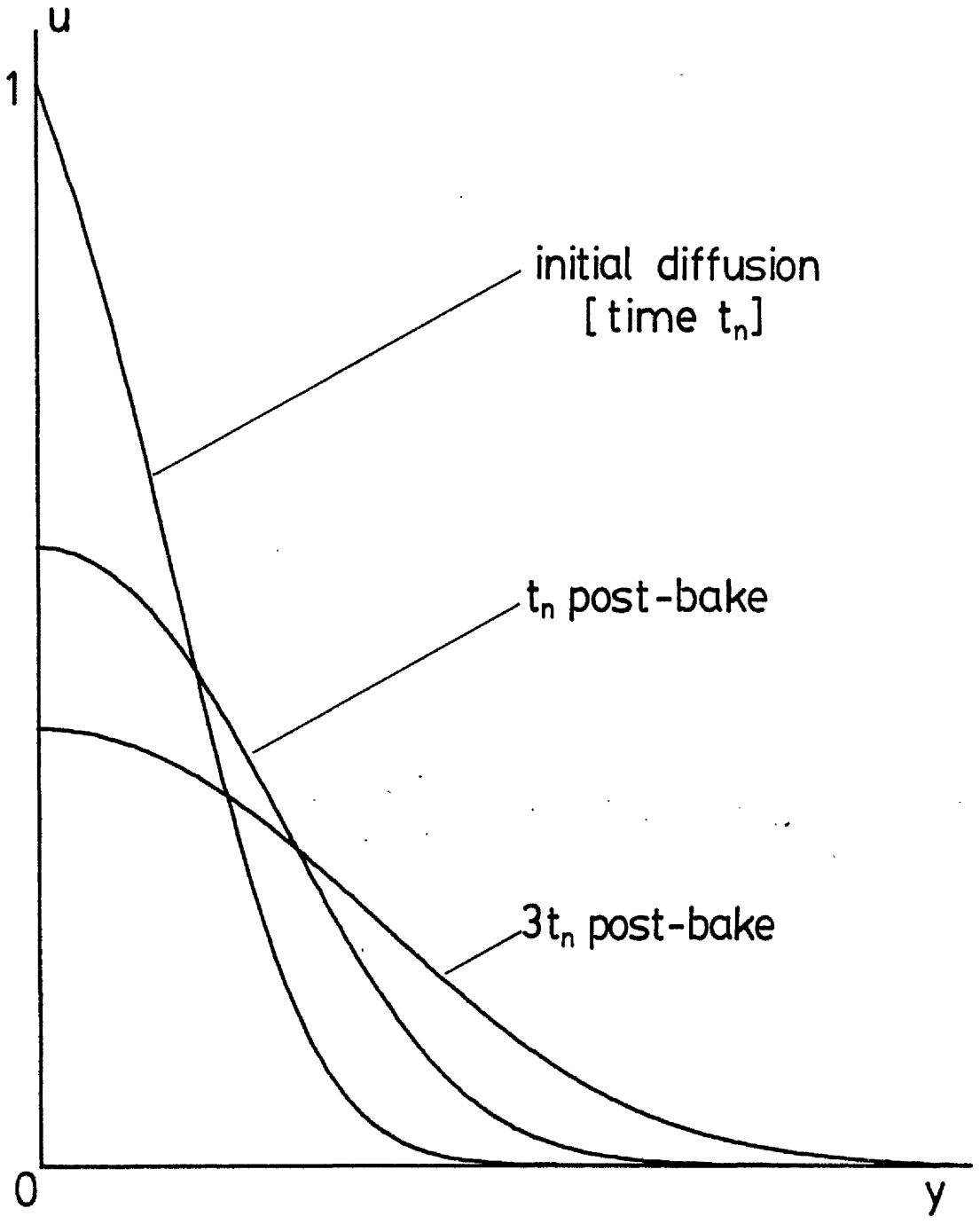
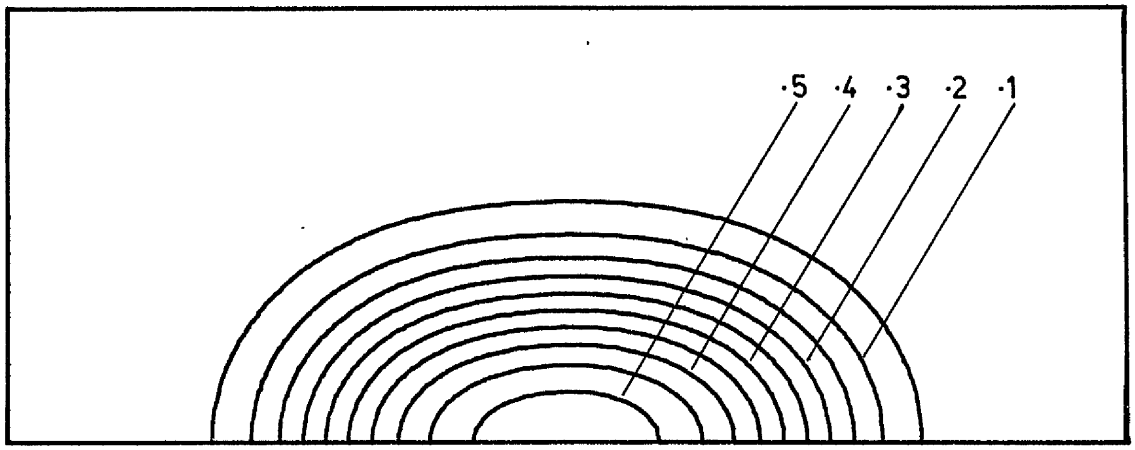


Fig. 5.5 Computed diffusion profiles for post-baked slab waveguides.

$\alpha = 0{\cdot}7$

(a)



(b)

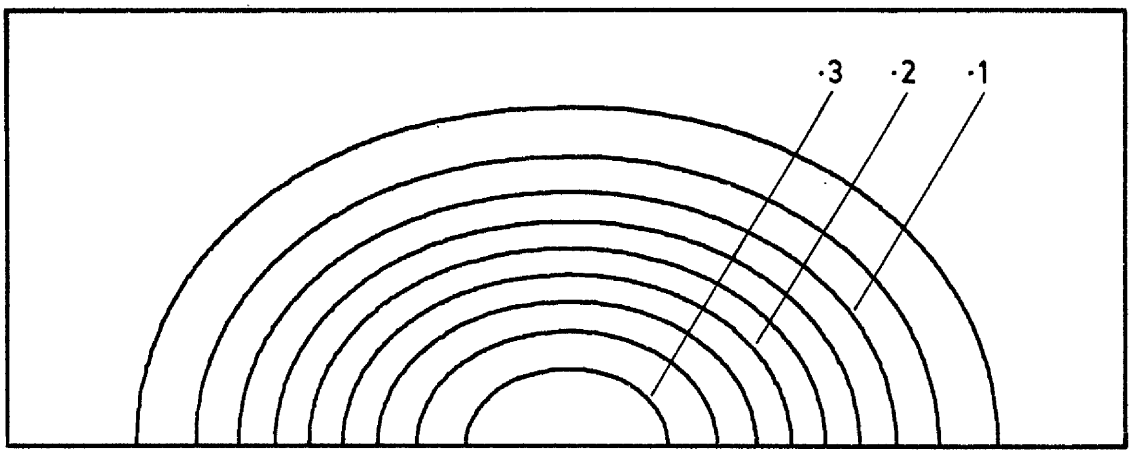


Fig. 5·6 The stripe diffusion of Fig.5·4(a) after a post-bake of: (a) t_n
(b) $3t_n$

[enlarged array, reduced scale]

e) Post-Diffusion Baking

The baking of a completed waveguide (discussed in section 3.6) is simulated by changing the source boundary condition (u = 1) to the reflecting boundary condition $\frac{\partial u}{\partial y} = 0$ and continuing the computation. The results of this process (which will, hereinafter, be referred to simply as "post-baking") are illustrated in Fig. 5.5, for slab guides, and in Fig. 5.6, for stripe guides. The post-baking times are t_n and $3t_n$ with $\alpha = 0.7$.

This distributed, limited source diffusion naturally results in a profile which tends towards a Gaussian in the slab case. In the stripe case, the concentration contours become more nearly semi-circular and the lateral diffusion is, initially, the more rapid.

5.4 Behaviour of the Wave Equation Solution

a) Convergence and Accuracy

Before the model can be used, it is essential to know how many terms are necessary for the series 5.2.4 to converge, with sufficient accuracy, to the true solution. Fig. 5.7 illustrates this convergence, for the slab case, by plotting the computed n_e against the number of basis-functions used by the 1-D variational program (SLAB). The refractive-index profile ($\alpha = 0.65$) has substrate and surface refractive incides of 1.5125 and 1.6035 respectively; the profile is scaled to give three slab modes.

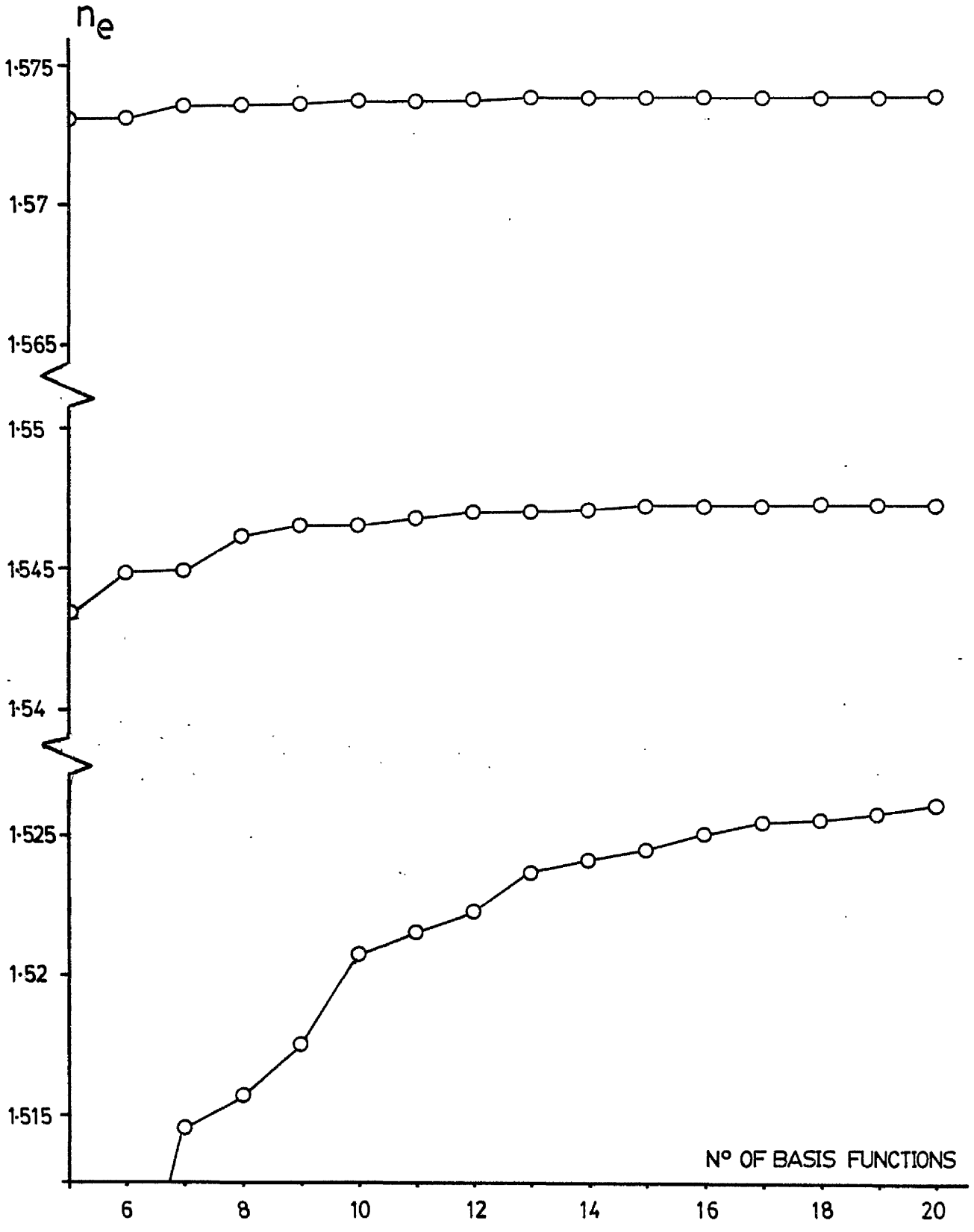


Fig. 5·7 Graph of n_e vs. number of basis functions, for slab waveguide

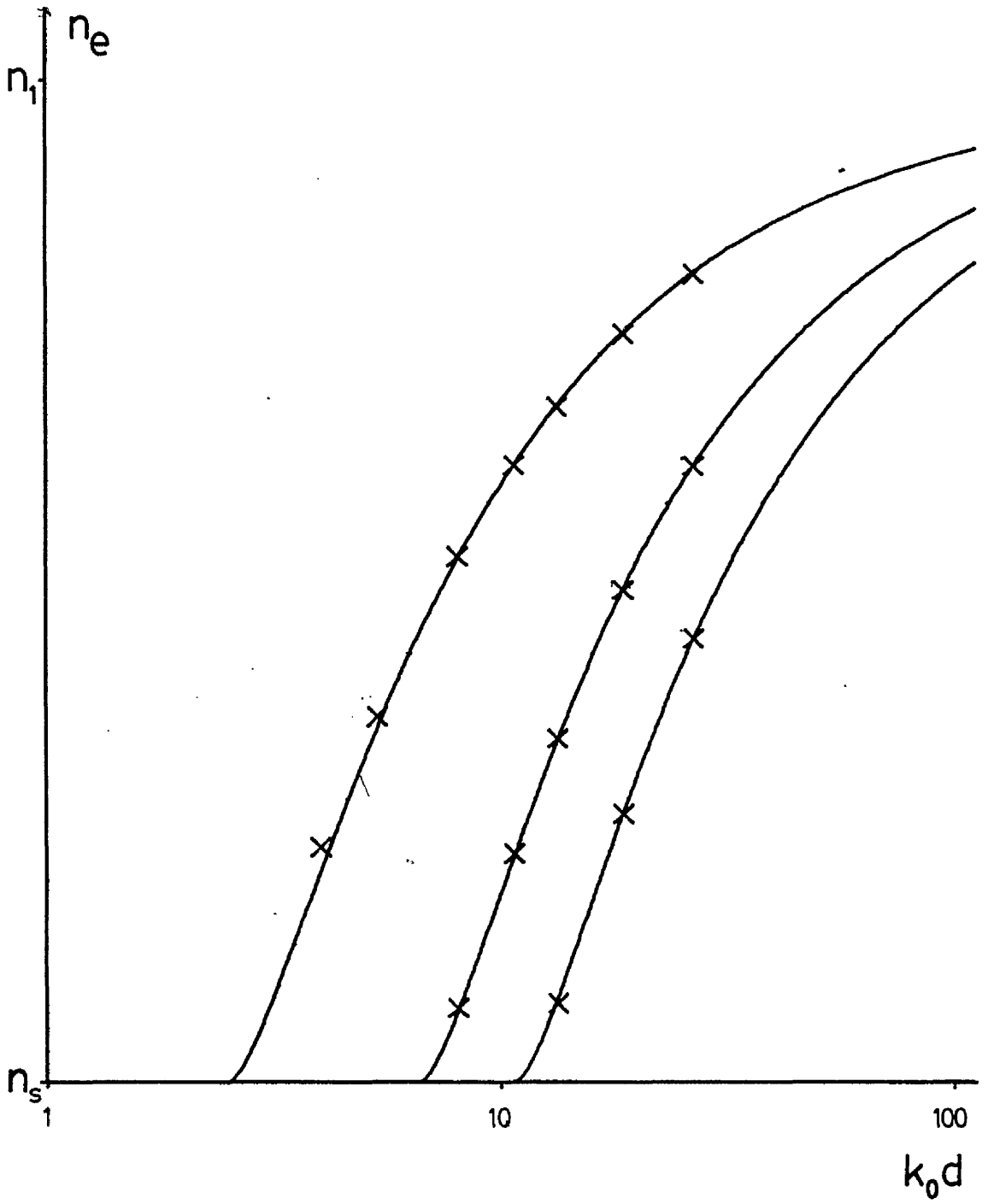


Fig. 5·8 Slab waveguide dispersion curves, comparing the W.K.B. method [——] with the Variational method [X]
($n_1 = 1{\cdot}605$; $n_s = 1{\cdot}5125$)

It will be noted that the fundamental mode converges for 10 or 13 basis functions (depending on the required accuracy - 10 is usually considered sufficient), whereas the second and third modes require 15 and 20 basis functions respectively. The basic shape of the field distribution for the nth mode is defined by the nth order basis function. It may be inferred, from the above, that the resemblance between this and the field distribution diminishes, as the mode order increases.

The variational method is compared with the W.K.B. method in Fig. 5.8. The variational solution (shown as points) uses 20 basis functions and the same profile ($\alpha = 0.8$, $t_n = 360$) is used by both techniques. As can be seen, the agreement is excellent in general, but error in the fundamental mode increases towards cut-off. It is not known which method is in error here - perhaps both.

b) Convergence in Stripe Guides

In order to obtain several lateral modes, a fairly wide, deep guide was used to test the convergence of WAVE. The parameters are:

aperture width: 7.1 microns

diffusion time: $t_n = 72$

$\alpha = 0.7$

surface refractive index: $n_1 = 1.605$

substrate refractive index: $n_s = 1.5125$

To test the effect of varying the number of x-direction basis functions, the number of y-direction functions was held at 10. The results are shown in Fig. 5.9.

A useful result is that a single x-direction basis function is able to model the fundamental mode with considerable accuracy. Clearly the best-fitting gaussian, which optimisation of ξ provides, resembles the actual field amplitude function fairly closely (see Fig. 5.13). In general, however, the resemblance of the nth order mode to the nth order parabolic cylinder function decreases as n increases. To model the first 6 modes (the maximum usually attempted in this work) 6 x-direction basis functions are necessary. To obtain the seventh mode, however, it is not clear that even 8 basis functions are sufficient. [Note that the 6th basis function here is D_{10} (symmetric mode) or D_{11} (antisymmetric mode)].

To test the effect of varying the number of y-direction basis functions, the number of x-direction functions was held at 6, the results being shown in Fig. 5.10.

Here, the lateral mode-spacing is almost invariant, revealing low coupling between the transverse resonances of the guide (see section 2.3). Most significant is the fact that only the parabolic cylinder functions D_0, D_3, D_6, D_9, D_{12} etc. appear to contribute significantly to the solution. Thus we use only 7, 10 or 13 basis functions in the y-direction when working with ion-exchanged guides. This must be due to the peculiar asymmetry of the y-direction field (see Fig. 5.13) which places limits upon its 'harmonic' content.

In general, the x-direction field, being either symmetric or anti-symmetric as are the basis functions, is more easily modelled than the asymmetric y-direction field and requires fewer basis functions to ensure accuracy.

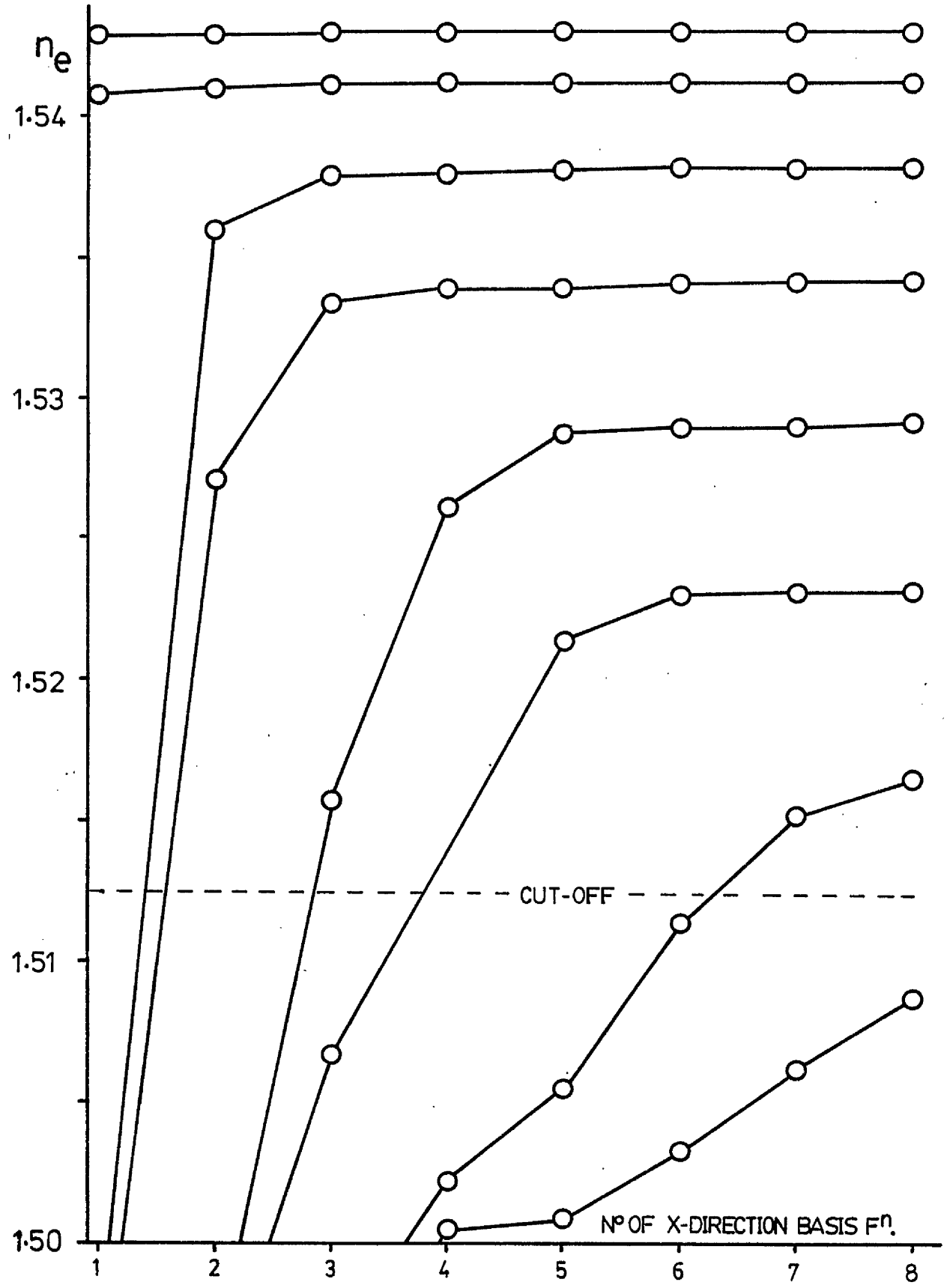


Fig. 5·9 Convergence of the variational solution [2-D] as the Nº. of x-direction basis functions is increased [lateral modes]

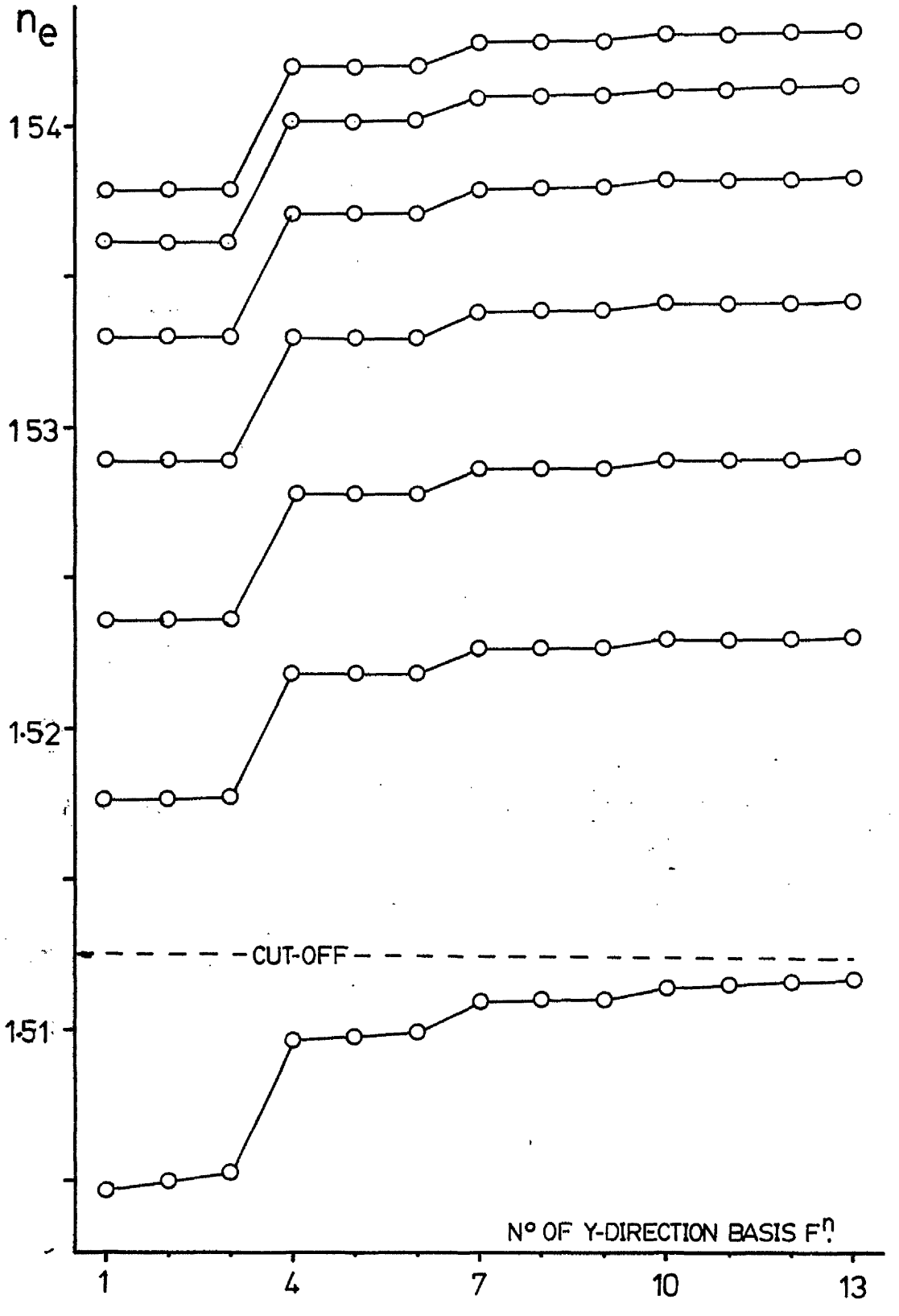


Fig. 5·10 As Fig. 5·9, but varying the N°. of y-direction basis functions.

c) Comparison of WAVE with Other Methods

WAVE has been compared with Goell's circular harmonic computer analysis[10] for rectangular waveguides. Goell gives his figures in terms of a parameter P^2 where:

$$P^2 = \frac{n_e^2 - 1}{n_1^2 - 1}$$

if the guide is surrounded by air; n_1 is the core refractive index.

The parameter B is 2.0, where:

$$B = \frac{2b}{\lambda_0} (n_1^2 - 1)^{\frac{1}{2}}$$

and b is the guide depth.

Taking $n_1 = 1.01$, we obtain $b = 4.463$ μm for $\lambda_0 = 0.6328$ μm.

The rectangular waveguide is approximated by an array, 81 rows deep whose elements are all of value n_1. 9 y-direction and 5 x-direction basis functions were used by WAVE. The results are compared with those taken from p. 2146 of (10) in the table below, for various guide aspect ratios, a/b, where a is the guide width.

a/b	P^2 (Goell)	P^2 (WAVE)
1	0.715	0.718
2	0.807	0.810
3	0.823	0.834
4	0.815	0.843

Note that when $P^2 = 0.715$, $n_e = 1.00716$.

As the width increases, P^2, as given by WAVE increases steadily towards the slab value as expected. Goell's figures do not behave so well however; the circular harmonic approach is only approximate and tends to break down when the aspect ratio is large.

WAVE has also been used to compute the modes of trapezoidal, ridge waveguides. This work, undertaken on behalf of C.S. Wilson, is reported in (61). It was noticed that the modes of these waveguides behave somewhat differently, as the number of basis functions is varied, from those of diffused guides. In particular, the lateral mode separation is not invariant with the number of y-direction functions, showing that the two transverse resonances are coupled as suggested in chapter 2.3 . Also more basis functions are needed to ensure accuracy. Nevertheless, agreement between WAVE and the finite-element method of Pelosi et al.[54] for trapezoidal ridge-guides is very good.

5.5 Dispersion Characteristics of the Stripe Ion-Exchanged Waveguide

Fig. 5.11 shows the computed effective index, n_e, of stripe ion-exchanged waveguides plotted against the diffusion aperture width. The slab effective index is 1.528 (n_1 = 1.605, n_s = 1.5125) which corresponds to about 11 minutes diffusion at 215°C. There is only one slab mode. Both of the simple boundary conditions (reflecting and u = 0) are used to model the diffusion mask, and compared with these is a dispersion characteristic calculated using the Effective Index method.

Clearly, the latter is inaccurate; increasingly so as the guides become narrower if the effective width is assumed to be that of the aperture. However Fig. 5.11 suggests that if an effective width, some 0.5-0.6 μm smaller than the diffusion aperture width were used, the Effective Index method would be more satisfactory.

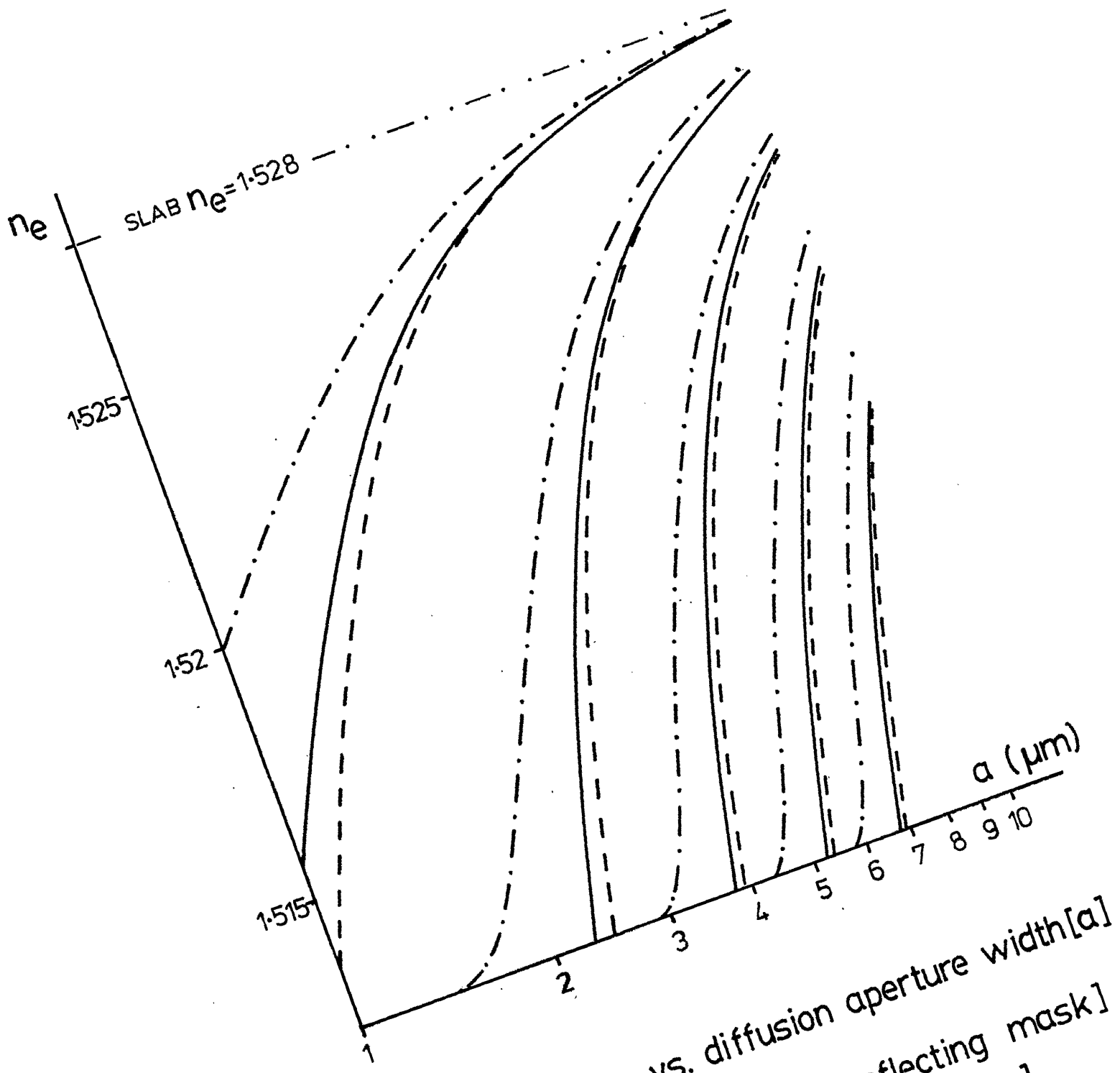


Fig. 5.11 n_e vs. diffusion aperture width[a]

——— WAVE [reflecting mask]

– – – – – WAVE [u=0 mask]

—·—·— Effective Index Method

The solution for the u = 0 boundary condition [see Fig. 5.4 (b)] is included mainly to demonstrate the self-correcting nature of the ion-exchange process. Even such a severe boundary perturbation as this, results in only a small reduction in effective width of between 0.1 and 0.2 μm.

WAVE and DIF become rather clumsy to use at greater than 10 μm width due to the large aspect ratio of the guide cross-section. In particular, it becomes difficult to optimise ξ,η and y_0 for wide and for shallow guides; N_{0000} has such a broad, flat maximum in such cases, that the program (whose accuracy is limited) cannot locate it.

The Field Amplitude Functions

The wave amplitude, $\psi(x,y)$ can be reconstructed from the eigen-vectors which WAVE computes. It may then be plotted either as a full contour map or as line plots, sampling the field amplitude, in either direction, through the point $(0, y_0)$.

Fig. 5.12 presents contour maps of the normalised field amplitudes for the three lateral modes of a 20 minute, stripe diffusion. The diffusion aperture width is 3 μm, and 10 y-direction and 6 x-direction basis functions were used. The contours shown are for $|E/E_{max}| = 0.1, 0.2 \ldots 0.9$ and the waveguide position is marked, in each case, by the $n(x,y) = n_e$ contour, beyond which the field is evanescent.

Fig. 5.13 shows line-plots of the field amplitude for the above case: the x-direction variation for the first two lateral modes (at $y = y_0$, $x > 0$ and using 10 x-direction basis functions) and the y-direction variation for the fundamental mode (at x = 0 and using 20 y-direction basis functions). Compared with each (shown dashed) is the basic, shape-defining basis function, obtained by stipulating that only one basis function be used in each direction.

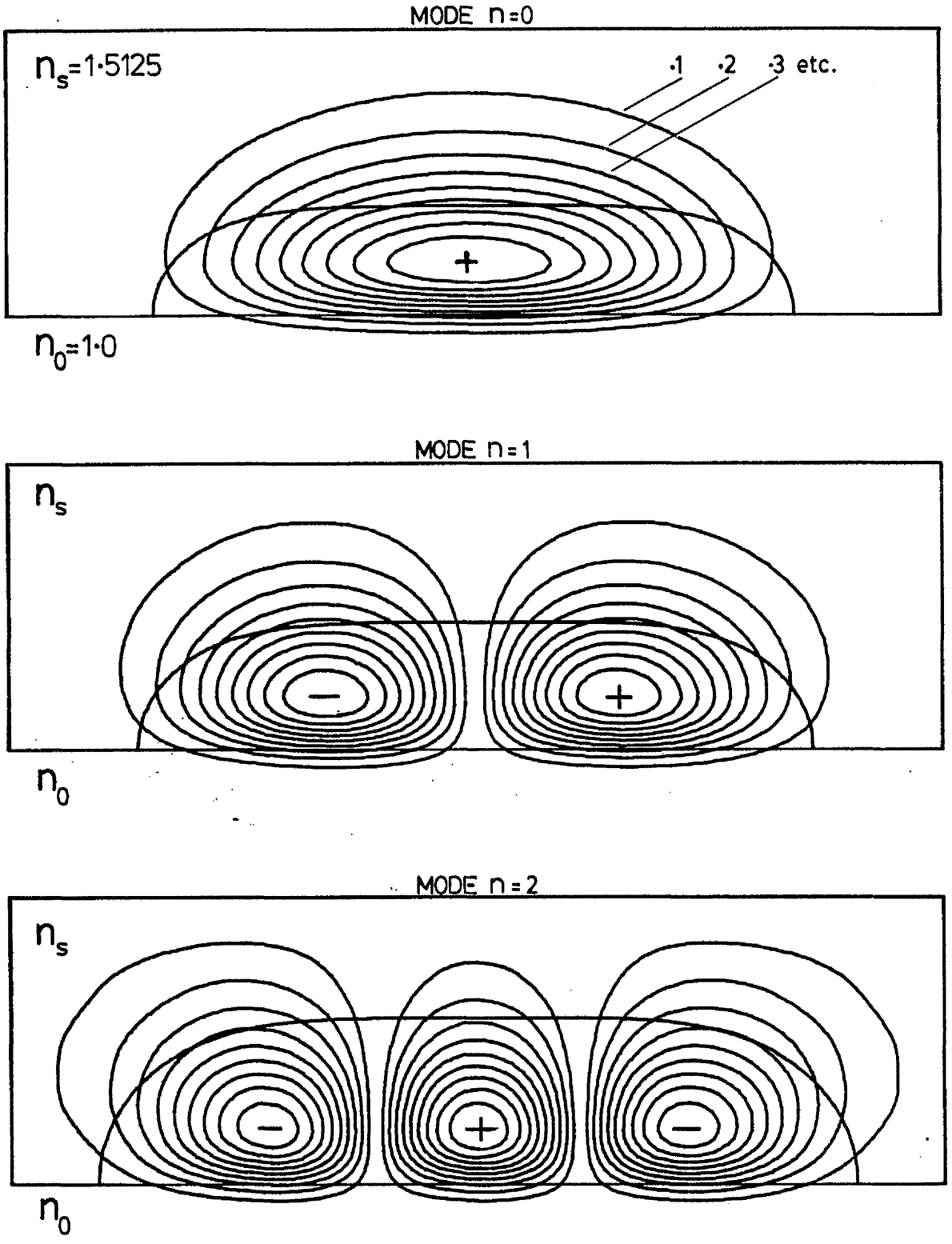


Fig. 5·12 Field amplitude distributions of a stripe, ion-exchanged waveguide [guide boundary at $n(x,y)=n_e$ contour]

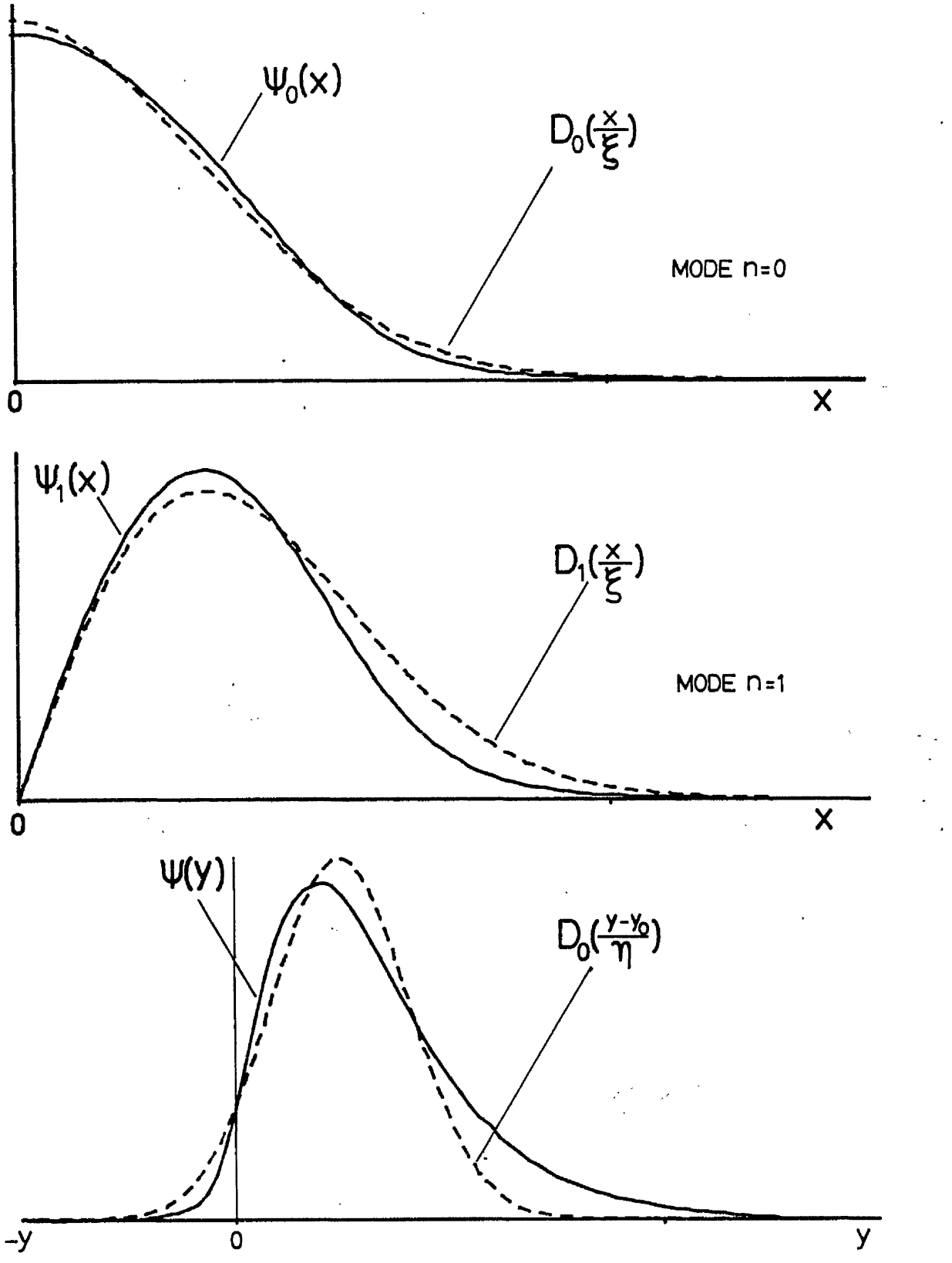


Fig. 5.13 As Fig. 5.12, but line-plots of the field amplitude[$\Psi(x,y)$] in x and y directions.

5.6 The Directional Coupler

To model a directional coupler, the whole of a waveguide diffusion profile is computed using DIF. WAVE, in reflecting this about the y-axis, interprets it as two waveguides side by side and solves it as a single, compound, guiding structure. The modes of such a structure have been considered in section 3.1.

The basic shape of the fundamental symmetric mode does not correspond to any single basis function, being, rather, a sum of D_0 and D_2. However, the basic shape of the first antisymmetric mode is clearly defined by D_1. Hence, to optimise ξ, η and y_0 for this computation, we maximise N_{1010} (see section 5.2) rather than N_{0000}. This gives a rather larger value for ξ but leaves η and y_0 unchanged.

The transfer length, L, is given by:

$$L = \frac{\lambda_0}{2(n_{e_s} - n_{e_a})} \qquad \text{(See 3.1.4)}$$

Since the small difference of two similar numbers is involved, the utmost accuracy is required. In practice, 12 x-direction basis functions (up to D_{23}) are found to be necessary since the fields of the coupler are more unlike the basis functions than are those of an isolated waveguide.

Figs. 5.14 and 5.15 show the fields of a single-mode directional coupler - both contour maps and line plots. The parameters are:

$n_1 = 1.599$, $n_s = 1.5125$, $\alpha = 0.7$

n_e (slab) = 1.5253

diffusion aperture width = 2.08 microns

guide separation = 1.493 microns

$n_{e_s} = 1.520800$; $n_{e_a} = 1.520699$

Therefore, at $\lambda_0 = 0.6328$ microns, L = 3.13 mm.

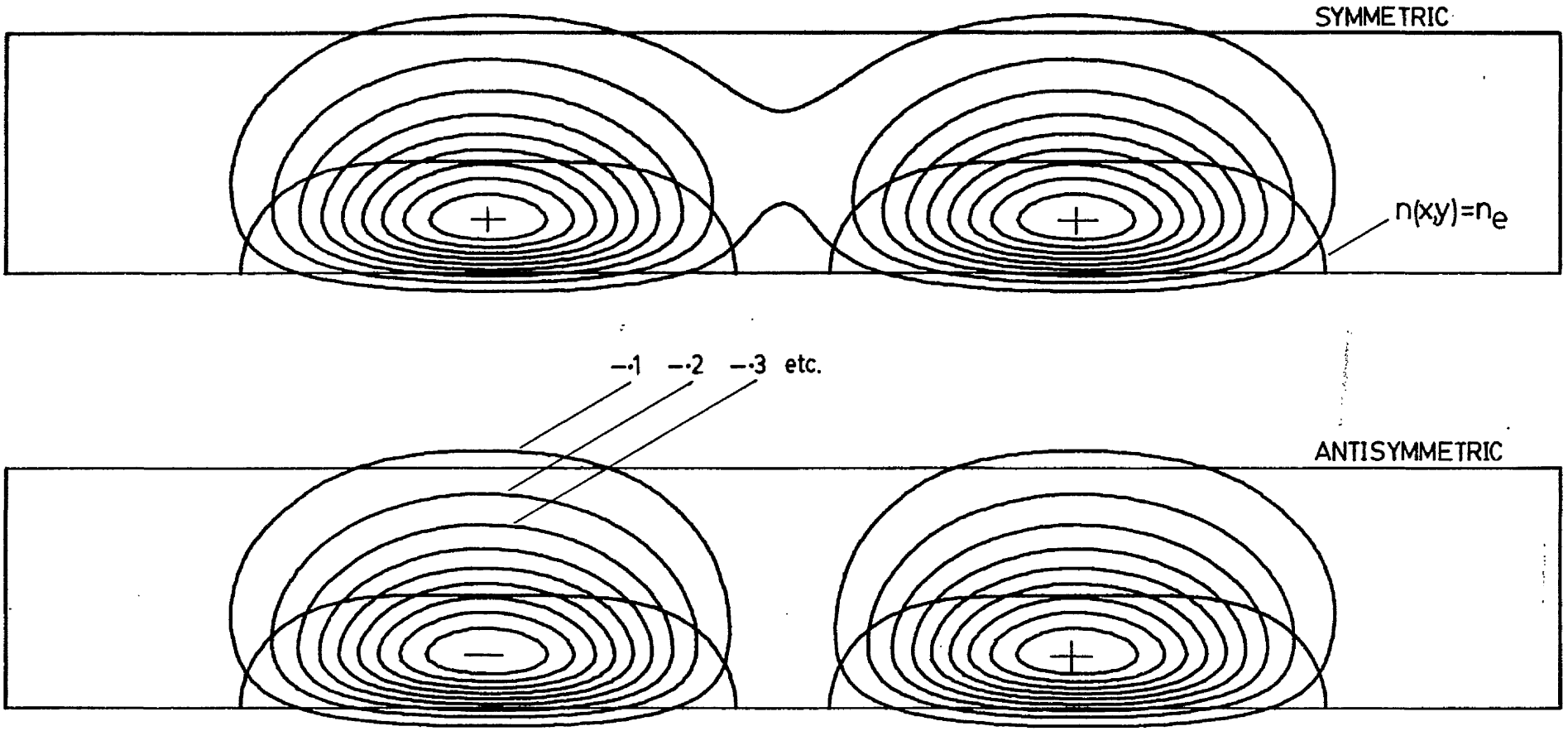


Fig.5·14 Directional coupler field-amplitude contour maps.[The frame indicates the size of the array U(x,y)]

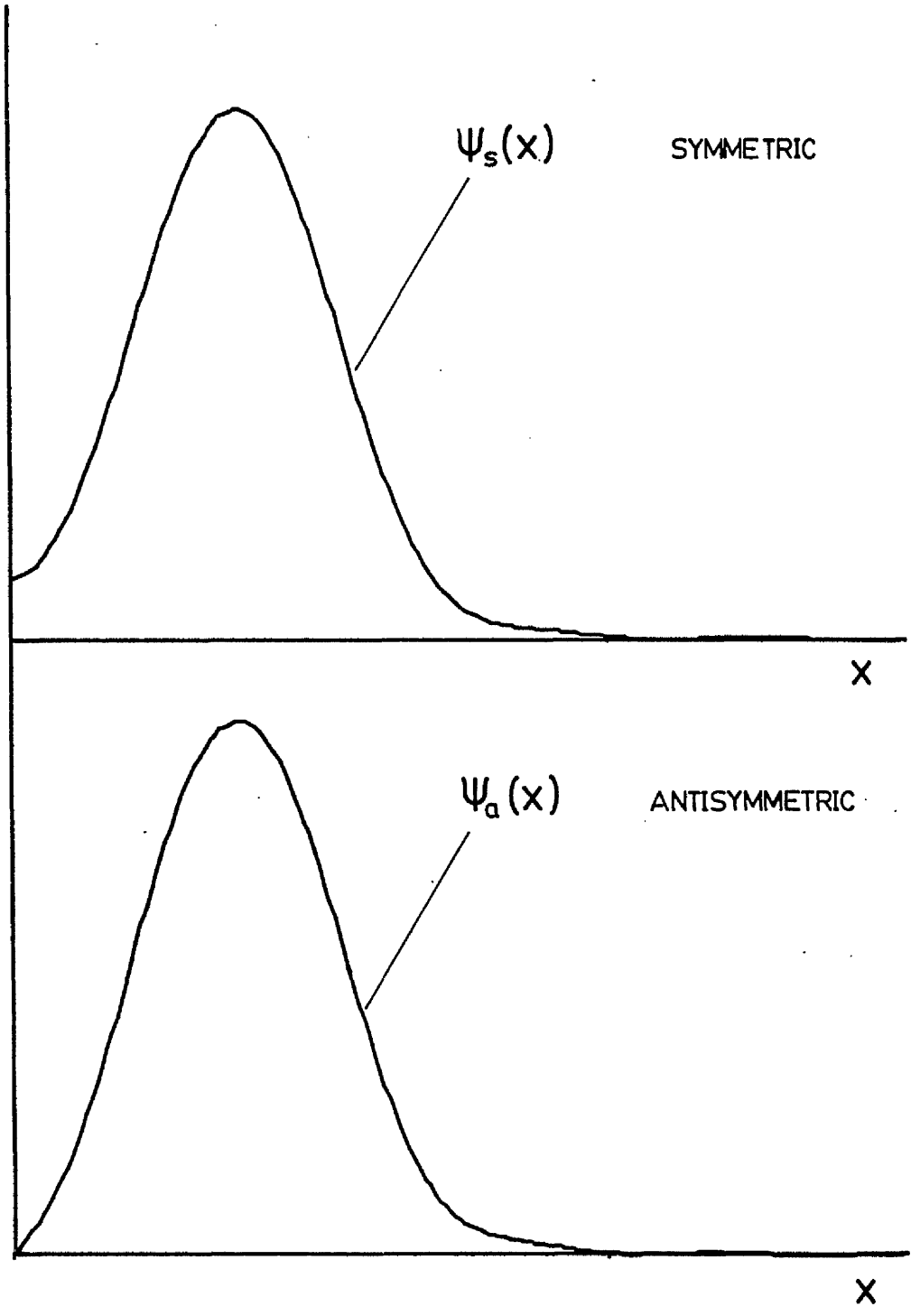


Fig. 5.15 x-direction field-amplitude plots of directional coupler modes.

PART III

CHAPTER 6

EXPERIMENTAL METHODS AND EQUIPMENT

6.1 Fabrication of the Aluminium Diffusion Mask

Standard photolithographic techniques are used to form an aluminium diffusion mask on the surface of a glass substrate (a microscope slide). There are many possible variations on two basic procedures, but all rely upon the properties of photo-resist. This is a photo-sensitive lacquer which is deposited on the substrate from solution by spinning and then drying.

The positive resists used in this work are degraded by short-wavelength light, the exposed regions becoming soluble in the alkaline developer solution. Thus, patterns of opaque and clear regions on an original shadow-mask may be transferred to the photo-resist film by contact printing. The resist film and shadow-mask are pressed face to face and illuminated through the mask. After developing, raised areas of resist corresponding to the opaque mask regions remain, the other areas being clear of resist.

The two basic procedures are as follows:

1) Chemical Etching

Photo-resist, deposited on top of the aluminium film, is exposed using a dark-field shadow-mask (i.e. the waveguide areas are clear) and developed. This clears away the resist, allowing the metal to be etched away where diffusion apertures are required.

2) Lift-Off

Photo-resist, deposited directly on the glass surface, is exposed using a light-field shadow-mask and developed to leave raised ridges of

resist where diffusion apertures are required. The aluminium film is then deposited. It adheres strongly to the clean, clear glass but does not extend up the vertical sides of the resist ridge. The unwanted metal is removed along with the resist by dissolving the latter in acetone.

The second method was used almost exclusively in this work (see Fig. 6.1). It offers good edge quality and width definition: also the areas to be ion-exchanged are protected from the developer solution and acid etches.

All fabrication procedures were carried out within a laminar air-flow tunnel of class 100 clean-room standard. This is situated within a larger area of nominally class 1000 clean-room standard. Filtered mains water and de-ionised water are available on tap: high purity (Analar grade) solvents were used, and disposable, vinyl gloves were worn at all times.

Further details of the fabrication procedure follow.

(i) Substrates

These are "Fisher Brand" microscope slides, which have an improved (fire-finished) surface quality compared with other brands.

(ii) Cleaning

This is a critically important stage; every trace of grease and other surface dirt must be removed, otherwise the aluminium film will not adhere to the glass surface. Some friction is found to be necessary initially, to remove gross contamination; therefore the slides are rubbed vigorously with sponge, detergent and water. Subsequent treatments

include rinsing and ultrasonic cleaning in water, acetone, and iso-propyl alcohol, and a long soak (up to 24 hours) in trichloroethylene to dissolve grease.

An important step is a soak in a proprietary, surface-active agent, "Decon 90". This alkaline soap digests organic contaminants and also possibly dissolves some glass from the surface. A weak, but hot, solution (3% at 90°C) of "Decon 90" is poured on to the slides and left to cool for several hours, which treatment is found to greatly improve the adhesion of the aluminium film.

Finally, the slides are thoroughly rinsed with de-ionised water, distilled water and finally, methanol. This latter is used for its hygroscopic properties, moisture being found to inhibit adhesion of the photo-resist to the glass. The methanol (ultra-pure, "electronic" grade) is blown off the slide with a jet of dry nitrogen and the photo-resist deposited immediately, before any dust settles on the surface.

(iii) The Photo-Resist

The positive working resist, "Shipley AZ1350 J" is used undiluted. This provides the fairly thick film (1.8 μm) necessary for the lift-off method. Line widths, smaller than 1.5 microns can be resolved in this resist. A few drops of the liquid resist solution are passed through a 1 μm filter on to the slide which is spun (on a vacuum chuck) at 4000 r.p.m. for 20 seconds, then dried at 80°C for 30 minutes.

(iv) Exposure

The contact printer is crude but effective, consisting of a lamp, above which the shadow-mask and substrate are pressed into intimate contact by means of weights.

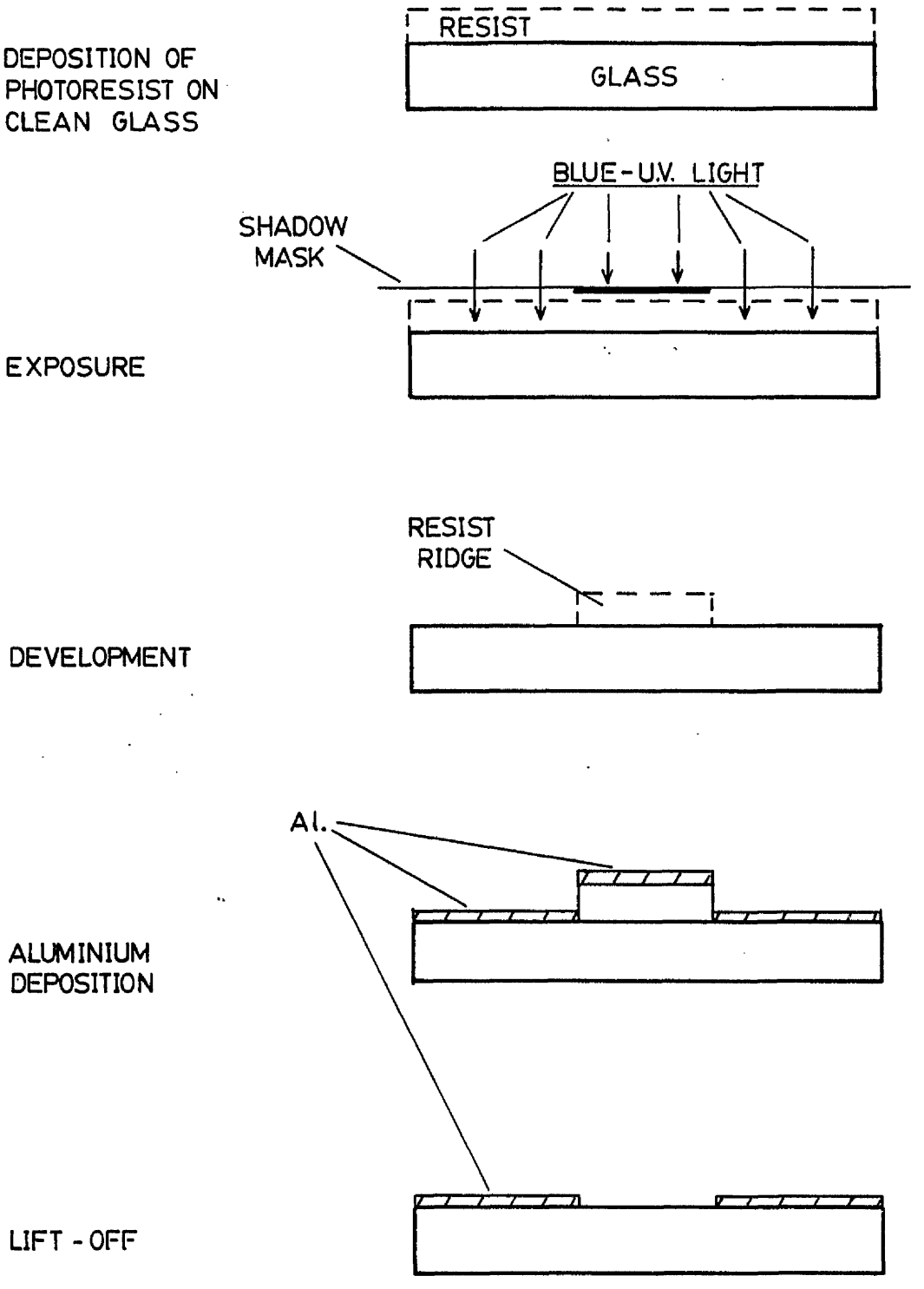


Fig. 6.1 Stages in the fabrication of an aluminium diffusion mask.

The shadow-mask is a glass plate, on which the pattern is defined in chromium. This is produced by photo-reduction of a large-scale art-work original, cut in "ruby-lith" film.

(v) Development

The usual developer solution is "Shipley AZ" diluted 1:1 with water. The minimum developing time is a function of temperature, exposure time and the condition of the developer (it is weakened by exposure to air). Since the developer also dissolves the unexposed regions slowly, over-developing reduces the line widths but is also bound to degrade the waveguide quality by emphasising width fluctuations. An alternative developer, "Shipley MF312", attacks the unexposed regions much more quickly and was occasionally used to reduce line widths.

The developing is halted by rinsing the slide thoroughly with de-ionised water. It is then blown dry and baked at 80°C for 30 minutes to dry it completely.

(vi) Aluminium Deposition

This is done by evaporation of the high purity metal from an electric heating filament under high vacuum ($\sim 3 \times 10^{-5}$ Torr.). The vacuum system is equipped with a rotary pump, for the initial rough evacuation, and a liquid nitrogen cooled, oil diffusion pump. The pressure may be reduced to 2×10^{-5} Torr. in about one hour's pumping.

The film thickness is controlled visually, by observing the white-hot aluminium source (at constant current) through the slides and interr-upting the deposition when the filament fades from view. This gives films of about 0.1 μm thickness, as measured by a stylus instrument (G.V. Planar Ltd.). The thickness is not found to be a critical parameter.

(vii) Lift-Off

The ultrasonic cleaner is used to shake off the surplus metal while the resist dissolves in acetone. The resist ridge is always slightly trapezoidal in cross-section and the aluminium is, therefore, occasionally continuous down its sides. This causes damage to the diffusion aperture during lift-off if the metal to glass adhesion is inadequate through poor cleaning.

It is useful to have such faulty samples thus revealed, since silver nitrate usually leaks under the mask at regions of poor adhesion [see Fig. 6.5 (b)] even when the mask is apparently perfect.

This completes the aluminium diffusion mask. If it is found to be sufficiently defect-free for use, the mask dimensions must next be measured. The measurement techniques are considered in section 6.7.

6.2 Anodisation of the Aluminium Diffusion Mask

Having measured the aluminium diffusion mask, we may either use it directly or render it electrically insulating by the process of anodisation.[58] This is an electro-chemical oxidation of the aluminium which is achieved by making it the anode of an electrolytic cell. The cathode may also be of aluminium and the electrolyte is dilute (20%) sulphuric acid.

When a voltage is applied between anode and cathode, sulphate ions (SO_4^{2-}) migrate to the anode and chemically combine with the aluminium to form a hard, strongly adherent oxide coating. With sulphuric acid

anodisation (unlike that of certain other acids) this coating is porous, so oxide growth is not interrupted by isolation of the metal surface. The process stops only when the glass surface is reached and all residual metal (if any) is isolated from the current source. The completely anodised film is transparent with a hint of grey (or blue) colouration.

The precise composition of the film is uncertain and depends upon the voltage, temperature and electrolyte concentration used. The porosity also depends upon these and tends to increase with temperature and concentration. However the composition is, approximately:

Al_2O_3 - 80%
$Al_2O_3.H_2O$ - 0.5%
$Al_2(SO_4)_3$ - 19%
H_2O - 0.5%

[from (58) p. 301]

The details of the anodisation process are by no means perfectly understood. Much space could be devoted to it. However we pause only to note the following:

(i) The pores in the oxide do not extend to the metal surface. There is a non-porous barrier layer, some 150 Å thick, between the metal surface and the porous oxide. It is not known which atom, aluminium or oxygen, diffuses through this barrier to maintain the oxidation. The pores are formed by hydration and dissolution of the barrier layer by the electrolyte and are about 100 Å in diameter.

(ii) The oxide can be sealed in hot water or steam. This converts some of the oxide to the mono-hydrate which, being less dense, swells and closes up the pores. This was tried, but was found to destroy the masking properties of the oxide film.

The Method

The temperature, during anodisation, must be kept below 20°C if the oxide is not to dissolve as quickly as it forms. It must be arranged, at the time of exposure, for the aluminium to extend to one end of the slide where the positive electrode is connected. The slide is immersed in the electrolyte as shown in Fig. 6.2.

It is found that the oxidation rate decays with distance from the electrode, with the result that the metal at the surface of the electrolyte anodises completely, cutting off the current while the lower regions are still metallic. For this reason, the current path must be protected by a film of photo-resist covering the entire slide except for a small "window" (opened by photolithography) where anodisation is required; this should be well below the surface.

It is advantageous for there to be a complete frame of protected metal surrounding the window, to ensure that no part of the area to be anodised is further from the effective current source than need be. Also, the pattern should be designed (or modified) to allow maximum current access to all regions of the mask.

The resulting oxide film will be quite even: this is further improved, however, if the electrolyte is very cold, thus increasing the metal conductance and reducing that of the electrolyte. A temperature of -13°C is used in practice; largely for convenience, the acid solution being kept permanently in a freezing cabinet.

It is important that the film anodise uniformly; otherwise some regions may be left with large amounts of residual metal, resulting in an electrical resistance which is insufficiently high. Good results with areas as large as 3 cm^2 have been achieved by the above techniques.

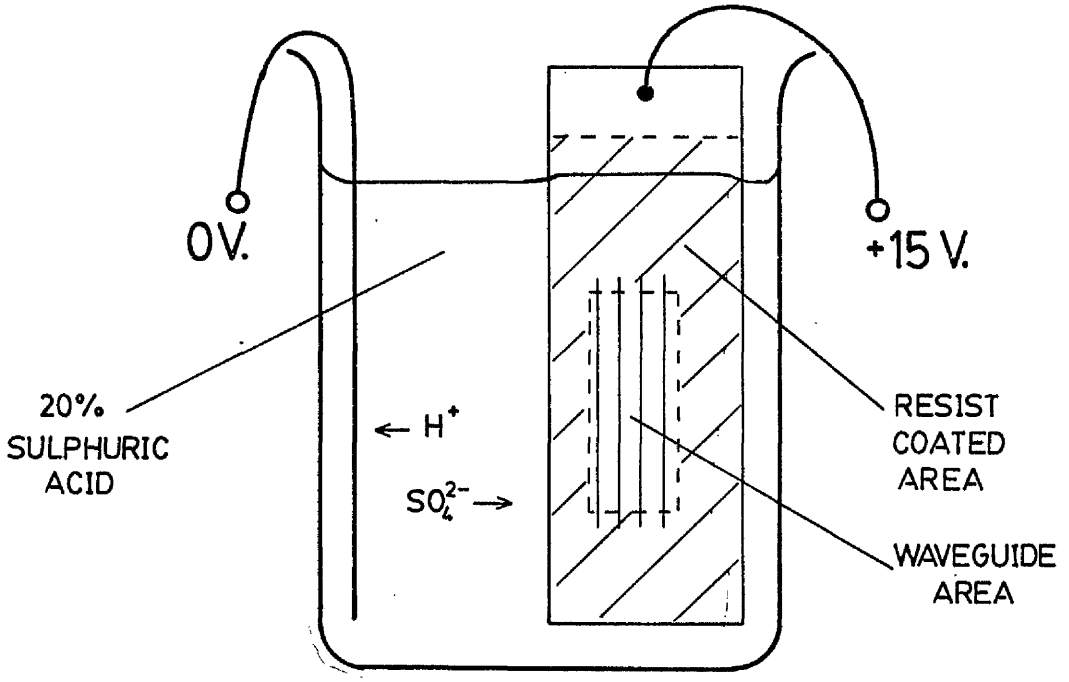


Fig. 6.2 Arrangement for the anodisation of an aluminium diffusion mask.

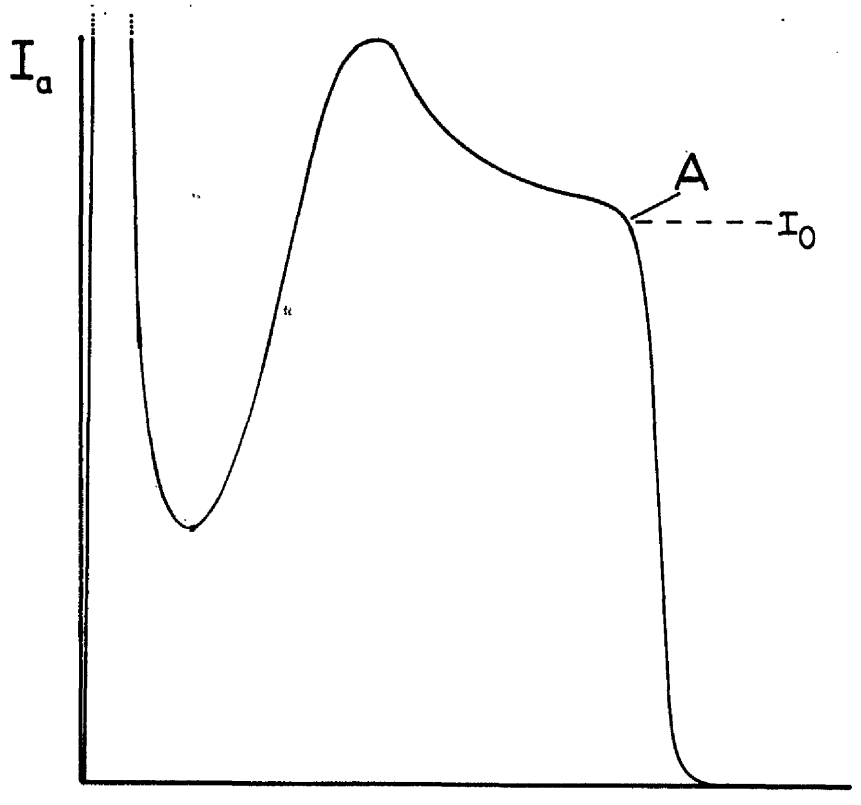


Fig. 6.3 Graph of anodisation current (I_a) vs. time.[constant voltage; thin Al. film]

Fig. 6.3 shows anode current plotted against time for a typical anodisation with a constant voltage of 15 V. The film is found to retain its metallic appearance up to the beginning of the final current drop (A), after which the continuity of the metal is rapidly lost. Latterly, the initial current spike was avoided by increasing the voltage gradually from zero.

Substrate cleanliness seems to be more critically important if the mask is to be anodised. A poorly adherent metal film may keep out the silver nitrate by virtue of its intrinsic cohesion; not so the anodised mask; it may even lose pieces during the anodisation. The use of hot "Decon 90" was introduced into the cleaning procedure to deal with this problem, and proved to be very successful; a good anodised mask can match a metallic mask at excluding the silver nitrate. Also, as will be seen (Chapter 8), it produces far superior waveguides. The natural porosity of the anodic coating need not cause any problems in this respect provided that the thin, non-porous, barrier layer is intact.

Clearly, a ring-resonator pattern could not be properly anodised, as the centre of the ring is isolated. This may not cause too great a problem, however (see Chapter 8 again). Alternatively, the fabrication technique outlined in Appendix 5 could be used; this has been tried and found to work, but has not been used extensively.

Another solution to this problem is to initially anodise a continuous film and then use the chemical etching technique (outlined earlier) to open the diffusion apertures.

After anodisation (or measurement, with metallic masks), the slides must be thoroughly cleaned, rinsed and dried. They are then ready to be diffused.

6.3 Diffusion

The diffusion equipment is similar to that described by Millar and Hutchins[45] and consists of a 5 cm diameter, horizontal tube furnace containing a semi-cylindrical bath of silver nitrate.

The microscope slide is supported by a cradle attached to a control-rod which passes through the tube furnace above, and to one side of, the silver nitrate bath. This horizontal arrangement (see Fig. 6.4) is designed to minimise temperature gradients. The control-rod may be moved through the furnace, to position the slide above the bath, and rotated, to immerse the slide for about two-thirds of its width.

The slide must be pre-heated prior to immersion; otherwise the silver nitrate freezes on to its surface, taking possibly twelve minutes to re-melt. Formerly, this was done by leaving the slide above the bath for from 10 to 15 minutes, but latterly, by pre-heating on a hot plate (hotter than the furnace) for 30 seconds before loading the slide into the furnace as quickly as possible. It is then left above the melt for about 2 minutes before immersing. The latter method saves time and is found (unlike the first method) not to disturb the temperature equilibrium within the furnace unduly.

The temperature control equipment is less elaborate than that described in (45). The control thermocouple (Iron-Constanton) which is placed in good thermal contact with the furnace-tube wall, is led directly to a 3-term controller (Control and Readout Ltd.). This has internal ambient compensation but may be set only in 1°C steps; however the zero-adjust potentiometer may be used to obtain continuous adjustment of the temperature setting.

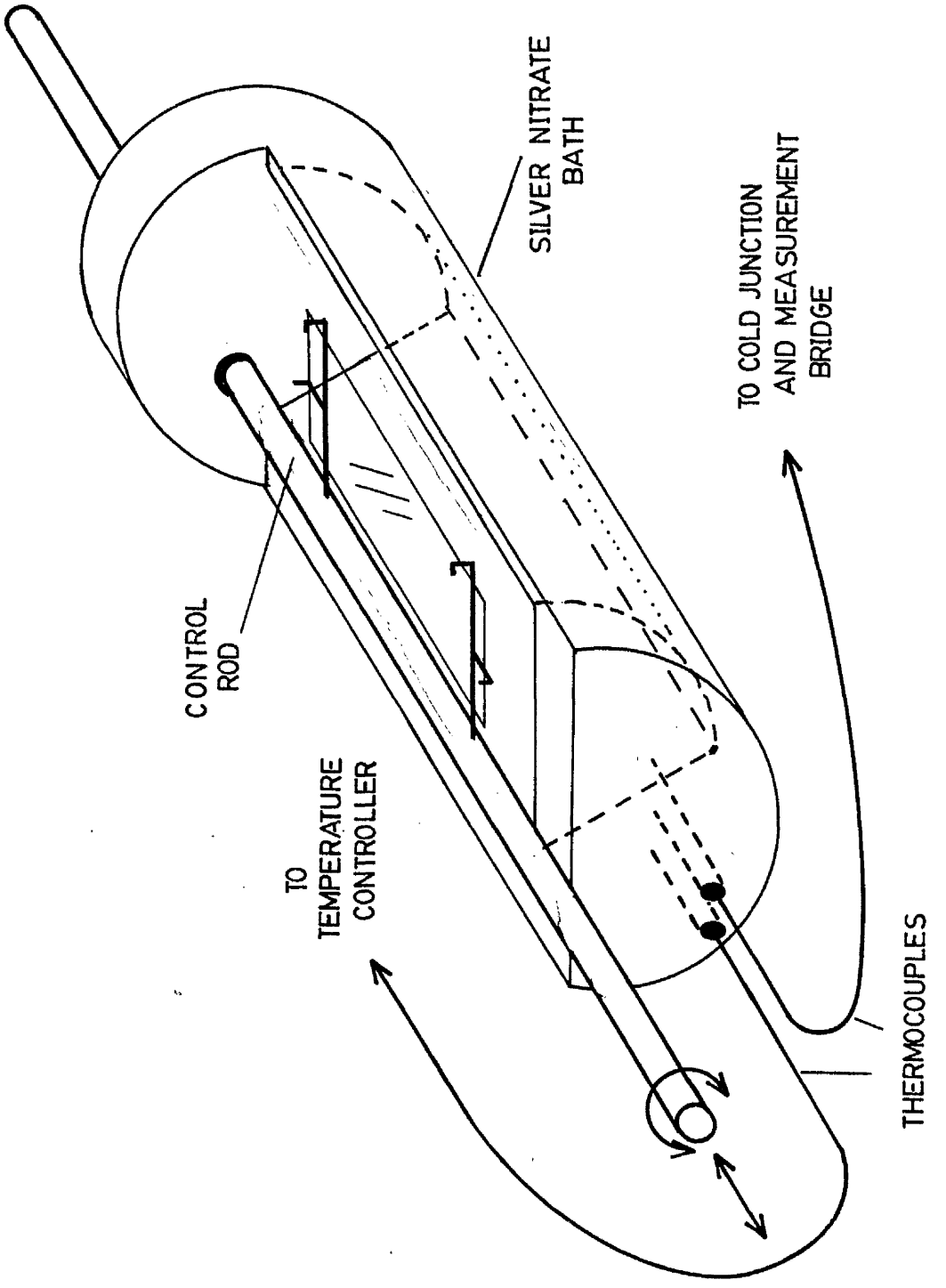


Fig. 6·4 The diffusion arrangement
[tube-furnace windings not shown]

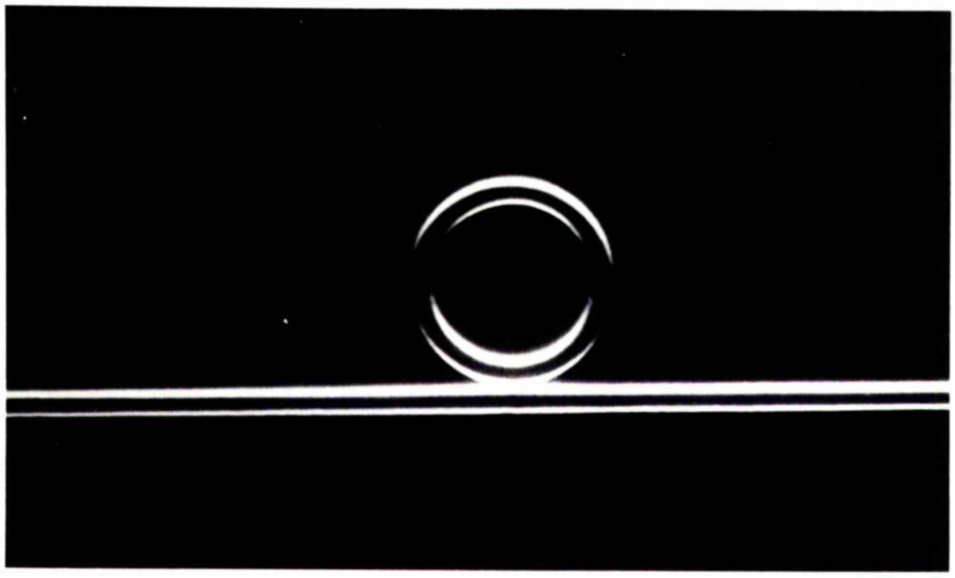
The measurement thermocouple is in intimate contact with the molten silver nitrate. This is led to a cold-junction at ice-point and thence to a standard-cell potentiometer. It is possible, with this system, to measure the temperature to one-tenth of a degree centigrade, but difficult to control it to that accuracy.

The best material for construction of the bath is a matter of uncertainty: fused silica and pyrex were both found to break if the silver nitrate freezes. Stainless steel was used for some time: the bath was machined from a solid piece of steel as the silver nitrate was found to corrode welds fairly rapidly. However, even this corroded slowly and polluted the melt. At present, a bath of high purity aluminium (99.99%), solid-machined and anodised (section 6.2), is used. This keeps the melt fairly clean, but the anodic coating must be renewed periodically.

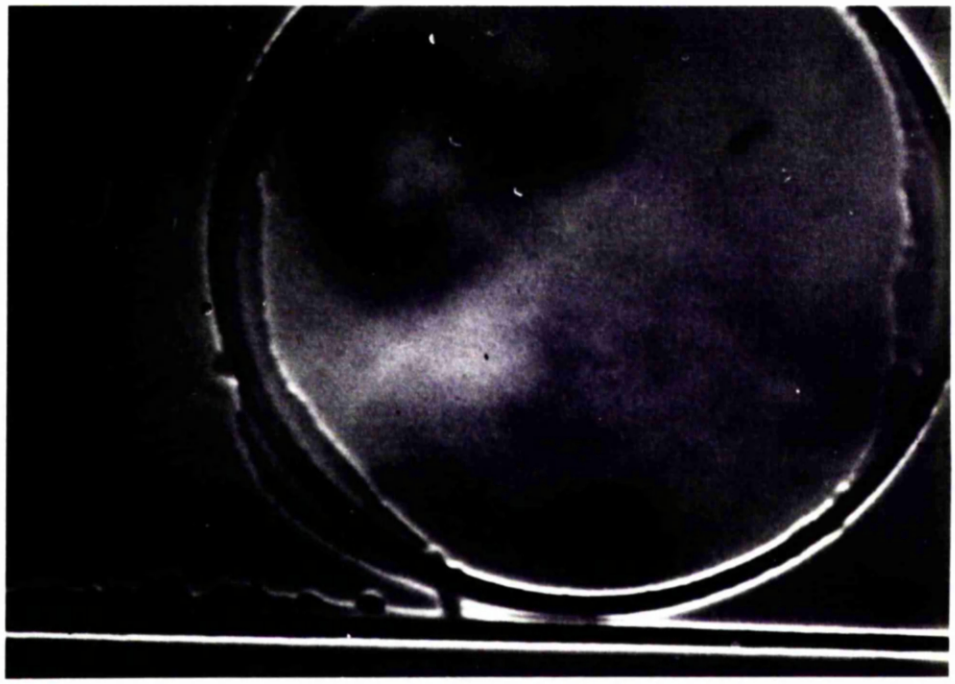
Silver nitrate is a powerful oxidising agent; it is very unstable when molten and tends to decompose spontaneously, aided by light and moisture. For this reason it is kept frozen when not in use - but sufficiently hot to prevent hygroscopic activity.

Completing the Waveguides

After diffusion the frozen silver nitrate is rinsed from the slides and the masks are removed using an etch which consists largely of orthophosphoric acid with 2% nitric acid and a little water. The anodised mask etches very much more slowly than the metal, so the etch is warmed (to about 40°C) in that case. Small areas of the mask material are usually retained to mark the waveguide ends.



(a)



(b)

Fig. 6.5 Phase-contrast photo-micrographs of ion-exchanged waveguides

(b) shows the result of poor mask adhesion.

The finished waveguides can be seen with the aid of a transmission, phase-contact microscope. This shows up refractive-index gradients as light, dark or coloured bands thus giving these buried guides the appearance of raised ridges.

Fig. 6.5 shows such phase-contrast micrographs. The pattern is that of an early (and abortive) attempt to make a ring-resonator. (It is now known that the rings were too small, and the coupling inadequate.) The diffusion aperture width, in this case, is 5 μm and the long diffusion time (2 hours) serves to show up the guides very clearly. The second photograph shows the leakage of silver nitrate which can occur if the aluminium adhesion is poor. These masks were not anodised.

6.4 The Post-Baking of Waveguides

If the baking process is to be accurately controlled and simulated, as outlined in section 5.3, the effective post-baking time must be known. This requires that the slide be brought up to the baking temperature very quickly, i.e. within 30 seconds. Two methods were considered:

(i) A block of metal within a furnace, slotted to accept the slide.

(ii) Immersion of the slide in a stirred bath of hot fluid.

It was decided that the latter alternative was the more acceptable.

The fluid used is "Dow Corning 550" silicone fluid. This high temperature oil is suitable for use at up to 230°C and has good heat transfer characteristics. The apparatus - illustrated in Fig. 6.6 - uses

a pyrex, 1 litre beaker, a 1 KW, coil immersion heater and a belt-driven propeller stirrer. The temperature control arrangement is the same as that used for the diffusion furnace. To make modelling simpler, the initial diffusion temperature is used, and only post-baking times which are integer multiples of the initial diffusion time are considered. If this multiple is kept constant, the final profile shape is, in theory, independent of the actual diffusion times involved. The stirring mechanism is not sufficiently robust to be left for long periods, so post-baking times of greater than one hour were not normally contemplated.

One effect of the use of silicone oil is that the glass surface becomes strongly hydrophobic. The oil itself dissolves readily in acetone but nothing seems able to restore the original wetting properties of the surface. Holland [(41) p. 335] comments on such chemisorbed silicone layers, noting that clean glass surfaces within a vacuum system which uses an oil diffusion pump (water cooled) become strongly water repellent. This silicone film is unlikely to be more than 30 molecules thick and similar films (see Holland p. 385) can be reduced to a mono-layer by rubbing. Thus, it will not affect the waveguiding properties.

A more serious effect is that, during post-baking, all the exchanged regions develop a yellow coating which is almost certainly an oxide (or other compound) of silver. The depth of colour of this coating increases with the diffusion/post-bake time and is dependent upon surface concentration; the second, and subsequent, post-bakes always cause much less yellowing than the first.

This coating causes no practical problems (e.g. increased loss) since it can be removed by wiping with a cloth or sponge. However, silver is clearly being lost from the system and this raises doubts as to the accuracy of the mathematical model.

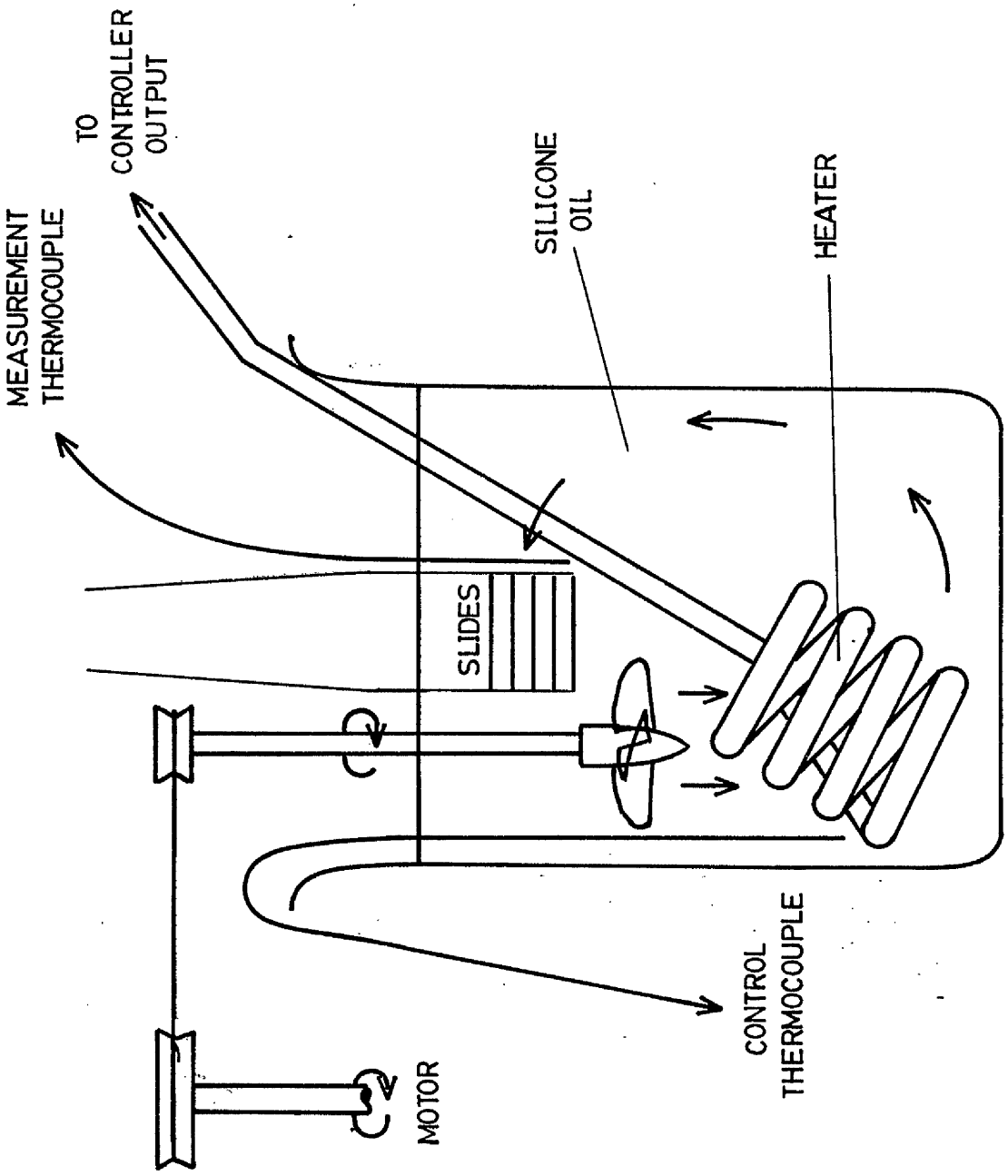


Fig. 6·6 Essentials of the post-baking apparatus

Similar brown coatings are developed by non post-baked guides during storage in dark, but atmospherically normal conditions. This also can be wiped off and indicates a high surface concentration of silver. Post-baking tends to obviate this problem.

6.5 Measurement of the Waveguide Modes

The waveguide modes are excited by means of a prism coupler[34] [see Fig. 6.7 (a)] which works on the same principle as the directional coupler (section 3.1).

The prism is clamped to the slide by means of a small G-clamp with a round-ended screw. This forces the waveguide surface into close proximity with the prism base over a small, approximately circular region which may be called the "coupling spot".

The input beam - usually from a He-Ne laser - is refracted by the sloping prism face and is incident on the prism base with an angle θ. The prism base and the guide surface are always separated by a thin layer of air, therefore, in general, if θ is sufficiently large, the input beam will be totally reflected from the prism base. (This emerges from the prism and may be cast on to a screen. The observed image will be called the "reflected spot".) Associated with this total reflection is an evanescent field impinging on, and propagating parallel to, the waveguide surface; its propagation constant, β_p, is:

$$\beta_p = k_0 n_p \sin \theta$$

where n_p is the prism refractive index.

This cannot be part of a guided field unless β_p coincides with β for one of the modes of the prism/waveguide coupled system, in which case:

$$\beta = \beta_p \qquad \text{and} \qquad n_e = n_p \sin \theta$$

i.e. the guide and the beam are phase-matched. This gives the waveguide effective index if the coupling is small.

When this condition is satisfied by the angle θ, total reflection is frustrated and energy flows from the beam into the waveguide. If the input beam is not parallel but has some angular spread, then only a part of the input energy will be absorbed; thus, with slab waveguides under certain conditions, a dark absorption line can be seen passing through the reflected spot and moving across it as θ is varied. When this line passes through the centre of the spot, it may be assumed that $\beta_p = \beta$, since the measured value of θ is that of the beam centre.

With narrow, stripe waveguides this absorption is too small to be seen and a different phenomenon must be used to identify phase-matching.

Of the light which passes into the waveguide, some immediately couples back out. This light has a lateral intensity variation appropriate to the mode launched and diffracts rapidly because of its own limited, lateral extent. It thus assumes the form of a long, narrow line, through the reflected spot, which is divided into lobes according to the lateral mode order. This is called the "input m-line"; its intensity passes through a maximum as θ passes the phase-matched condition, which may be thus identified. An input m-line can often be seen with slab guides also, but this requires a coupling spot small enough to diffract the light. Slab m-lines are never lobed.

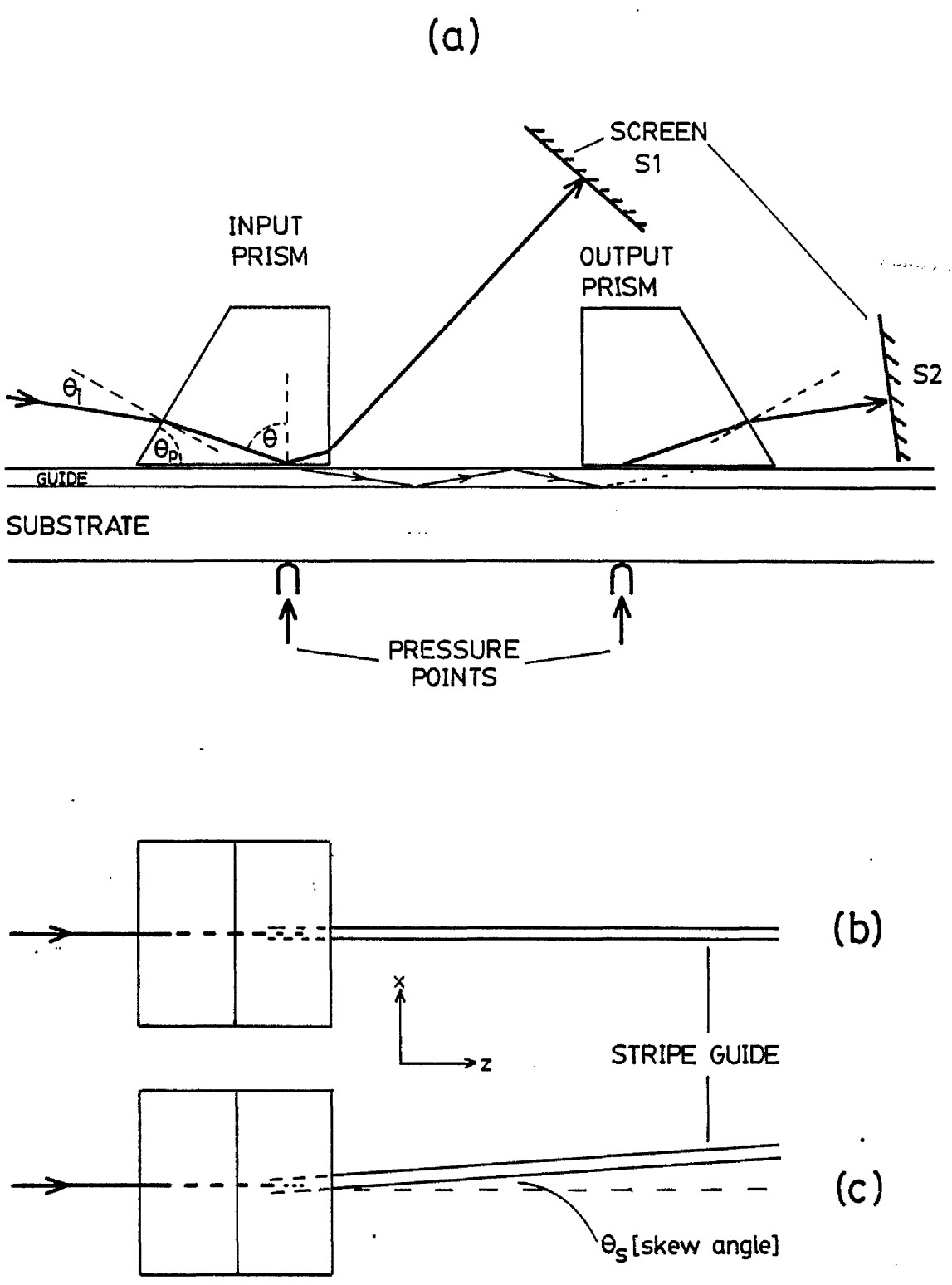


Fig. 6.7 The Prism Coupler

(a) side view. Input m-lines are seen on screen S1 and output m-lines on S2

(b) and (c) top view; arrangement (b) excludes odd-order modes. These require $\theta_s \neq 0$ (c).

Because of this out-coupling, the efficiency of the prism coupler is limited to about 80%, with rather less than this being practically achievable.[34] The energy in the guide may be maximised by positioning the input beam near the right-angled corner of the prism so that the coupling is interrupted at a suitable point.

Light may be extracted from the waveguide by means of a second prism further along [see Fig. 6.7 (a)]. This coupling is one-way and the efficiency is limited only by the size of the coupling spot. The out-coupled light takes the form of output m-lines, identical to the input m-lines but "cleaner"; i.e. lacking the scattered and reflected light from the input beam. They emerge at the same angle as the input beam enters; thus, either input or output conditions may be used to identify the coupling angles and hence, the effective indices (n_e) for each mode.

From Snell's Law and the prism geometry:

$$\theta = \theta_p + \sin^{-1}\left[\frac{\sin \theta_i}{n_p}\right]$$

$$\text{and } n_e = n_p \sin \theta \qquad (6.5.1)$$

where θ_i is the incident angle of the external beam to the sloping face of the prism, and θ_p is the prism angle.

Strong coupling tends to reduce the angular selectivity of the waveguide by damping the transverse resonance: it also changes n_e alightly. Thus, with both methods, the coupling should be minimised for the sake of accuracy.

In this work, the input m-lines were used exclusively, for measurement purposes. The reasons are as follows:

1) Measurement of the output coupling angles requires extra equipment and difficult alignment.

2) A bright output m-line spectrum requires a large guided power density. This is much more difficult to achieve than a bright input m-line.

3) Using the input method, any mode which can be excited can be measured. This includes leaky modes which are somewhat below cut-off. The output m-line spectrum reveals only those modes which propagate as far as the output prism which, in a lossy guide, may exclude high order modes. Only the input m-line method may be used with short or damaged waveguides.

The two methods were compared for slab waveguide modes and were found to be in good agreement as regards the values of n_e. Only the input m-line method will be considered in detail.

The Stripe Antisymmetric Modes

Equation 6.5.1 assumes that the stripe waveguide, the input beam and the prism-face normal all lie in the same y-z plane [see Fig. 6.7 (b)]. However, only the even-order (n = 0, 2, 4 . . . .) lateral modes may be excited in this way. The antisymmetric (odd-order) modes require some phase-shift across the guide, which can be supplied by rotating the guide about the y-axis, with respect to the input beam [see Fig. 6.7 (c)].

The formula 6.5.1 may be generalised to include this skew-angle, θ_s:

$$n_e = n_p \sin \theta . \cos \theta_s$$

where θ is given by 6.5.1 (a).

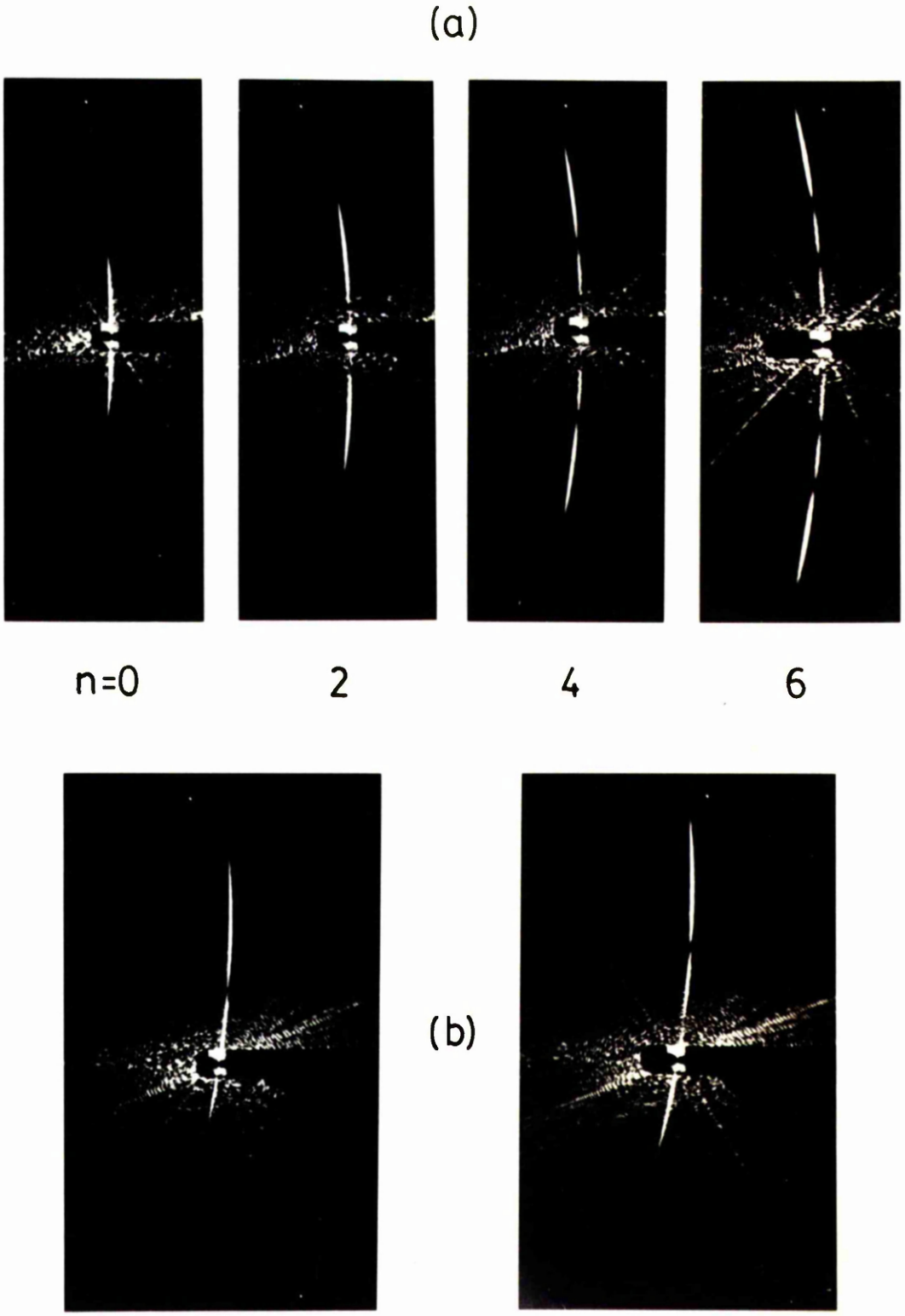


Fig. 6.8 Input m-lines. (n is the lateral mode order)

(a) Zero skew-angle

(b) Non-zero skew-angle

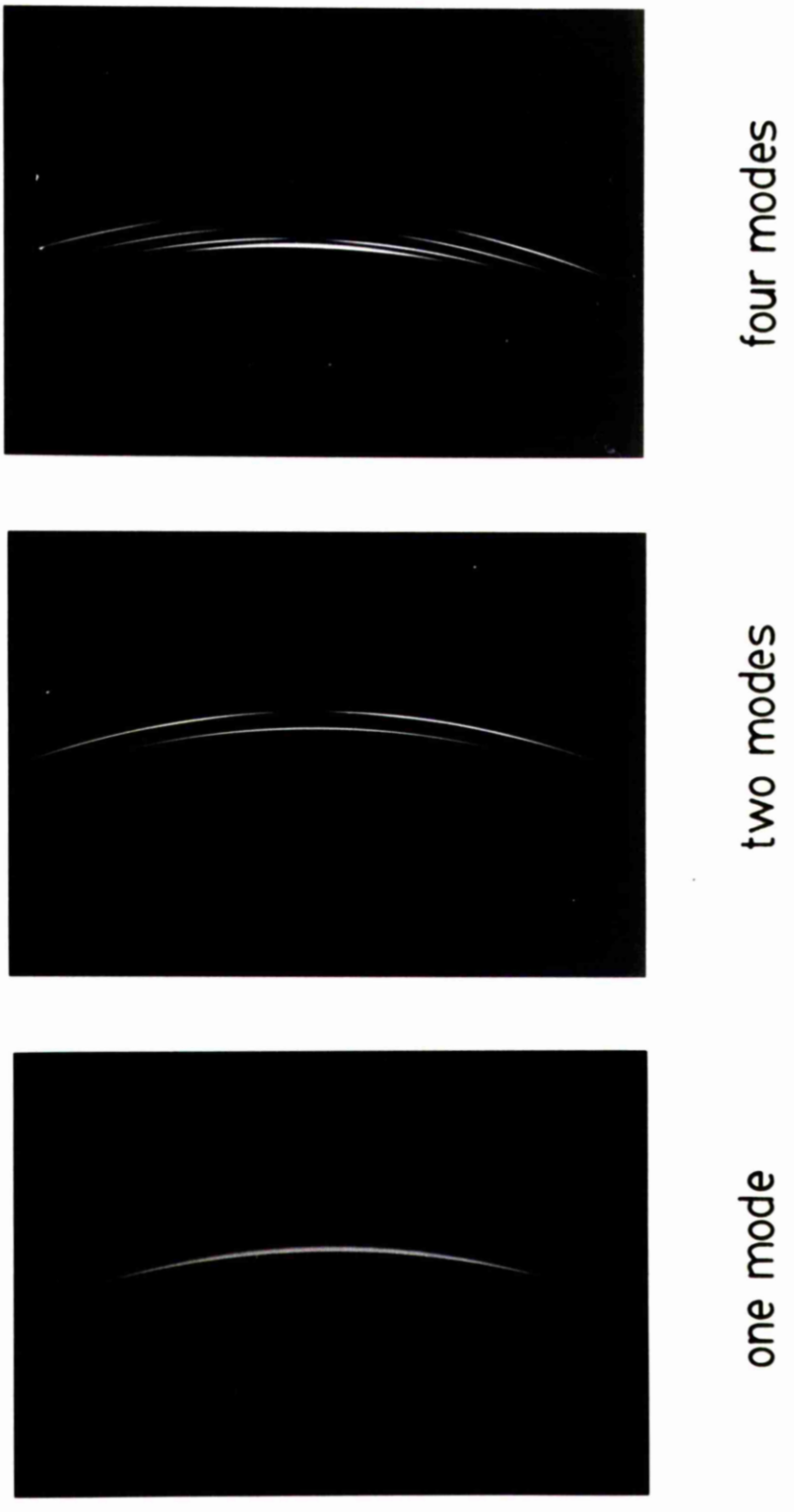


Fig. 6.9 Output m-lines of three ion-exchanged waveguides which support the stated numbers of modes.

In general, if the reflected spot coincides with a zero of the m-line pattern then the mode cannot be excited.

Fig. 6.8 (a) shows photographs of input m-lines for the four symmetric modes of an 8 μm wide waveguide. In each case the reflected spot (blanked out) masks the central maximum of the pattern. Fig. 6.8 (b) shows the first antisymmetric mode (n = 1) and also the mode n = 2 with a skew angle to reveal the central maximum.

Fig. 6.9 shows output m-line spectra for stripe guides which support one, two and four modes. In this case, light was launched into the stripes, via a tapered section, from a slab region, so all the modes are excited simultaneously. As with direct excitation, some asymmetry in the launching conditions helps to excite the odd-order modes.

6.6 The Optical-Bench Arrangement

The slide is mounted in a jig, providing several degrees of freedom. Translation in three dimensions is available, as is rotation about each of the x, y and z axes. The slide is positioned with the x-axis (defined in Chapter 2) vertical; however, the coordinate system need not be re-defined on this account.

The azimuthal rotation (about the vertical x-axis) is the most important, since it is this which is used to locate and measure the coupling angles. Positioning and measurement to 0.01° is possible by means of a digital, angular read-out device (Microcontrol Ltd., France).

A linearly polarised helium-neon laser (λ_0 = 0.6328 μm) with a nominal, maximum output power of 5 mW, is used to excite the waveguides. Only T.E. (E vertical) polarisation is considered.

The optical lay-out used, (Fig. 6.10) offers a compromise to several conflicting requirements. In order to efficiently excite narrow stripe waveguides, several of which may be close together, a fairly small beam diameter is required. The long path between laser and lens allows any given lens to produce a smaller focal spot than would a short path. Also, the long path makes possible an accurate measurement of the angle between the input beam and the normal to a reflective surface. The zero angle is that for which the reflected beam is colinear with the input beam, as defined by an aperture through which it passes. The coupling angles and the prism angles θ_p (and also n_p) were measured in this way.

For mode measurements, the beam should not be too tightly focussed since the beam convergence adds to the uncertainty in the coupling angle. The use of a lens with a focal length of 0.5 m, placed 2 m from the laser and 65 cm from the prism is found to provide a suitable compromise. This gives a spot diameter of 150 μm [according to gaussian beam theory - see for example (3) Chapter 6] as compared with 680 μm at the laser. The aperture, for location of the zero input angle, may be placed after this lens (as is desirable); a distance of 65 cm is found to be adequate to give this zero-setting to an accuracy of ± 0.01°.

It is essential that the input beam remain at the same point on the prism base as the angle θ_i is varied. For this to be possible, the beam must pass through the centre of rotation. Thus, alignment of the system is rather critical.

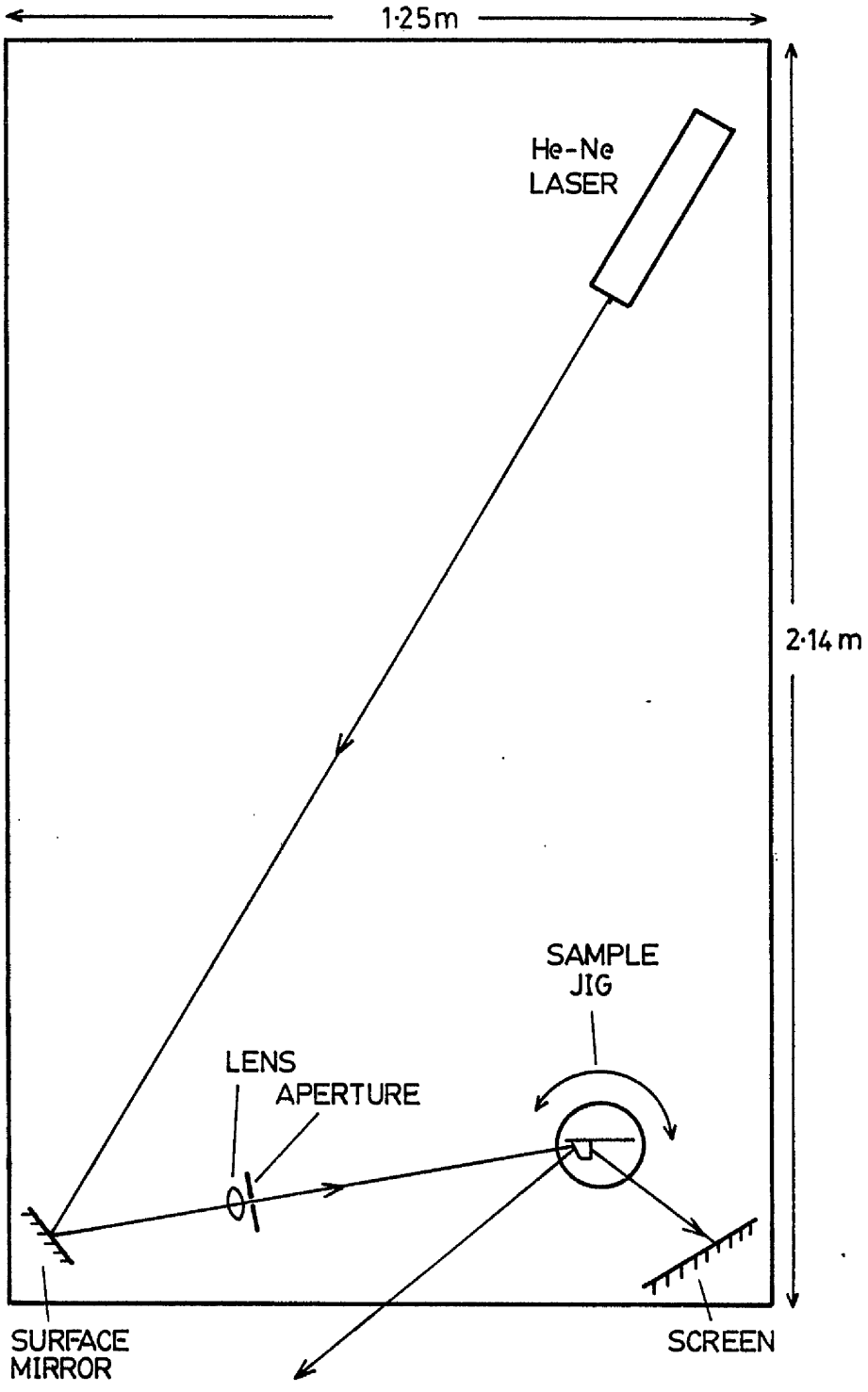


Fig. 6·10 Optical bench layout

Summary of the Measurement Method

1) Slab Waveguides

Coupling may be adjusted by means of both the prism-clamp screw and the Z-translation carriage (by moving the coupling spot away from the beam). The dark absorption line is located, and the coupling is reduced to the minimum for which it can still be clearly seen. The angular read-out device is then zeroed at the angle for which the reflected beam is colinear with the input beam: the angle for which the absorption line passes through the centre of the emerging beam is then noted.

2) Stripe Waveguides

As above, but the m-line is used instead of an absorption line. Minimum coupling is used and the angle of maximum m-line intensity is noted, taking the median of several attempts to locate it. The zero skew-angle is found by adjusting this until no odd-order modes can be excited. This angle may be very small, in which case great accuracy is not required.

The total measurement accuracy is estimated at $\pm$ 0.025°.

6.7 Measurement of Diffusion Aperture Widths

It is common practice to use an optical microscope, equipped with either a filar (cross-wire) or an image-shearing eyepiece, to measure sub-millimeter line-widths. The weaknesses of such methods are discussed in reference (57); they are as follows.

(i) The object itself is not measured, but only its optical image. If reflected light is used, a diffusion aperture is a dark band against a bright background; if transmitted light, it is a bright band against a dark background. Because of diffraction effects, the widths of these images (assuming perfect focus) differ by about 0.3 μm for a 5 μm line.

(ii) The eye-pieces used are calibrated against a standard graticule. However, the centre-to-centre spacing of graticule markers 1 μm apart bears little relation to the image width of a 1 μm wide diffusion aperture.

(iii) The image width is critically dependent upon the focus setting, the judgement of which is unreliable. Uncertainty of edge location and aberrations of the optical system itself also add to the error.

Single Slit Diffraction

Measurements made using an image-shearing eye-piece were found to be so unreliable that an alternative was sought in single slit diffraction. Since the aluminium diffusion mask for a straight, stripe waveguide is an almost ideal diffraction slit, its width may be deduced from the far-field (Fraunhofer) diffraction pattern. The wavelength of He-Ne laser light is accurately known and may, therefore, be used as the reference length.

When a slit of width, a, is illuminated by a plane, monochromatic light wave, the intensity of the far-field diffraction pattern on a planar screen, (at distance, d, from the slit) varies as:

$$I \propto \left[\frac{\sin \alpha}{\alpha}\right]^2 \qquad \text{as} \qquad \frac{d}{a} \rightarrow \infty$$

where $\alpha = \frac{\pi a}{\lambda} \sin \theta$

θ is the angle between the centre line and the observed point at (d,x). Thus, $\text{Tan } \theta = \frac{x}{d}$.

The derivation of this can be found in any basic physics text [e.g. (62)].

The nth intensity minimum occurs for $\alpha = n\pi$, thus, having found θ for the nth minimum, a may be calculated.

$$a = \frac{n\lambda}{\sin \theta} \qquad (n \triangleq 1, 2, 3 \ldots) \qquad (6.7.1)$$

The Practical Arrangement

The equipment is basically the same as that used for measuring the waveguide modes (section 6.6).

A long focal-length lens provides a beam neck of several hundred microns diameter at the specimen; thus the beam divergence should not affect the measurement significantly. The screen, which may slide on a length of optical bench, is coloured red but is calibrated vertically (x) in white. Low level, blue, ambient light illuminates the scale, but is not reflected by the red screen which remains dark. Thus the red (λ = 632.8 nm) diffraction pattern and the centimeter scale can both be seen clearly, without interference.

The screen is moved until the chosen intensity zero coincides with a suitable scale marking; x and l may then be measured, allowing tan θ, and hence, a, to be calculated.

Accuracy

The method is highly self-consistent; the same width may be measured many times using the same, or different diffraction minima without varying by more than $\pm$ 0.04 μm. A more serious source of error is the actual width fluctuations of the aperture. With apertures made very narrow (< 2 μm) by overdeveloping, the pattern has minima rather than zeros and the measurement is an average over a length equal to the spot diameter.

Double Slit Interference

When the device to be made is a directional coupler, the diffusion mask is a double slit, with slit width, a, centre-to-centre spacing, b, and inner edge separation, c. The intensity variation is given by the product of a diffraction function and an interference function.[62]

$$I \propto \left[\frac{\sin \alpha}{\alpha}\right]^2 \cos^2 (\delta) \qquad (6.7.2)$$

where $\delta = \frac{\pi b}{\lambda} \sin \theta$

Thus, for the mth interference minimum:

$$b = \frac{(m + \frac{1}{2})\lambda}{\sin \theta} \qquad (m = 0, 1, 2 \ldots) \qquad (6.7.3)$$

a is given by 6.7.1.

The main difficulty with this measurement, is identification of the distinct diffraction and interference minima. For the dimensions of interest in this work, b is rather less than $2a$, so the minima for $m = 1$ and $n = 1$ almost coincide. b may be obtained accurately using the first minimum, ($m = 0$) from which the precise position of the interference minimum $m = 1$, may be inferred; it will be at one side of the broad minimum which results from the near coincidence. This enables the position of the first diffraction minimum ($n = 1$) to be estimated (at the other side of the broad minimum) with little error.

Fig. 6.11 is a computer graph of the diffraction pattern for typical coupler dimensions, using 6.7.2. This small program was written to aid interpretation of the observed patterns.

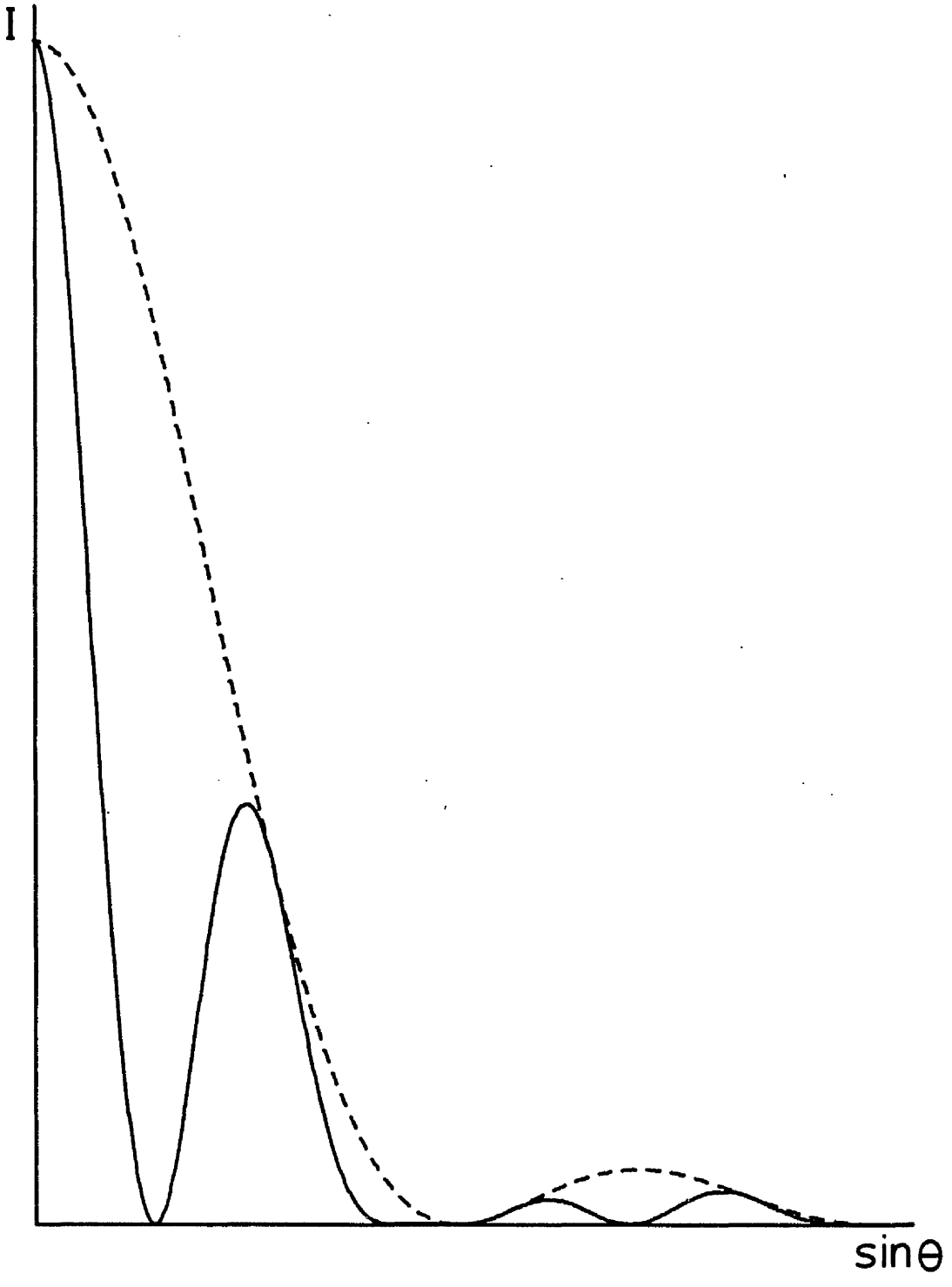


Fig. 6.11 Intensity variation, $I(\sin\theta)$ for single-slit [- - - - -] and double-slit diffraction [——]

$\lambda = 633\,nm$; $a = 2{\cdot}2\,\mu m$; $b = 3{\cdot}9\,\mu m$

6.8 The Prism Angle and Refractive Index, and the Substrate Refractive Index

The coupling prisms are cut from "Schott SF-15" glass which has a refractive index of 1.69425 ± 0.0002 at a wavelength of 632.8 nm. The prism angles, which are about 60°, were measured to the nearest 0.01°, as suggested in section 6.6, by locating the normals to each face. The refractive index was also measured by finding the critical angle for total reflection from the glass to air interface. This gave a value of 1.6936 ± 0.0002 with respect to air, or 1.6941 with respect to a vacuum. This is well within the quoted tolerance, so the value $n_p = 1.6936$ is used.

The substrates used are "Fisher-Brand" microscope slides whose approximate composition is given in Appendix 2. Their refractive index has been measured using an abbé refractometer (newly calibrated) and by measuring the cut-off angle in waveguides. When $n_e = n_s$ substrate modes appear; these are easily identified, but there is, inevitably, some spread in the critical angle.

n_s is found to vary from slide to slide, but it usually lies between 1.512 and 1.513. In this work, n_s is taken to be 1.5125.

C H A P T E R 7

ION EXCHANGED SLAB WAVEGUIDES

Before the mathematical model can be used to full effect it is essential to determine the parameters of the diffusion process: the peak (surface) refractive index, n_1, the self diffusion coefficient of silver ions in glass, D_a, and the diffusion non-linearity constant α ($\alpha = 1 - D_a/D_b$ where D_b is the self diffusion coefficient of sodium). This is done by adjusting these until the best fit between measured and computed values of n_e is found.

This matching procedure is most conveniently carried out using slab waveguides, since the distribution of the slab modes gives the best indication of the profile shape as determined by α. As has been shown (section 5.4) SLAB can compute only the 3 lowest order slab modes accurately, unless more than 20 basis functions are used (which is impractical). Thus, it is necessary (and more convenient) to use the W.K.B. method.

7.1 Parameter Values Given by the Literature

The melting point of pure silver nitrate is at about 210°C. In order to avoid problems of colloidal silver formation (yellowing of the glass), which is sometimes caused by exchanging at high temperatures, and to obtain sufficient control over the diffusion depth at very short diffusion times (single-mode guides are required) the comparatively low temperature of 215°C was used throughout the work. However, in all the literature in which parameter values are suggested, much higher temperatures - greater than 300°C are used.

Doremus,[50] using a radio-active tracing technique, obtains the following self diffusion coefficients:

sodium ; $D_b = 1.02 \times 10^{-13}$ m^2/sec

silver ; $D_a = 8.5 \times 10^{-15}$ m^2/sec

at 374°C

Therefore $\alpha = 0.9167$ at 374°C.

Unfortunately, this value cannot be used in the present work due to the temperature difference. In another reference [(40) p. 36] Doremus states:

"There are very few data on the variation of the mobility ratio with temperature. The mobility ratio, μ_b/μ_a, of sodium to silver ions in a soda-lime glass decreases appreciably from 375°C to 320°C. . . . . These results all imply that the slower ion . . . . has a smaller activation energy than the faster ion."

If the ratio μ_b/μ_a decreases as temperature decreases, then it follows that, at 215°C, α will be somewhat less than 0.9167.

Stewart et al.[44] consider the problem of the diffusion profile in some detail, though again rather high temperatures are used throughout (> 280°C). Using a mode-matching technique, a parabolic curve is found to provide a good fit to the measured mode spectra. The 1-D diffusion equation may be solved as an infinite power series. When the first three terms of this are compared with the best-fitting parabolic profile, a value of $\alpha = 0.56$ is obtained. As shown in Fig. 8 of (44), this value gives a theoretical profile which matches the parabolic profile only over its upper half ($u > 0.5$). The implication is that the real profile has a considerably attenuated "tail" region (i.e. that part of the profile where $u \longrightarrow 0$ assymptotically) compared with the theoretical model. Also, it suggests that there may be no value of α for which the computed profile exactly matches the experimental data.

The surface refractive index (n_1) is found, by Stewart et al., to lie between 1.6 and 1.605 for Chance-Pilkington microscope slides. In this work, Fisher-Brand slides were used for their superior surface quality; thus, n_1 may differ somewhat from that quoted in (44).

The value of the diffusion coefficient, D_a, at 215°C, may be calculated from that given by Doremus using the Arrhenius equation:

$$D_a = D_0 \exp(-\Delta H/RT)$$

where D_0 is a constant which includes the entropy term:

$\exp(\Delta S/R)$.

ΔH is the Activation Energy

R is the universal gas constant

T is the absolute temperature.

Stewart[44] obtains a value for ΔH of 8.5×10^4 J/mole. Assuming that, at 374°C, $D_a = 8.5 \times 10^{-15}$ m^2/sec, then at 215°C:

$D_0 = 6.02 \times 10^{-8}$ and $D_a = 4.83 \times 10^{-17}$ m^2/sec

Doremus[50] uses a value for ΔH of 20 Kcal/mole (= 8.37×10^4 J/mole). In this case, at 215°C:

$D_0 = 4.73 \times 10^{-8}$ and $D_a = 5.23 \times 10^{-17}$ m^2/sec.

Unlike n_1 and α, D_a is rarely used explicitly in this work; it is more convenient to work in terms of the normalised diffusion time (t_n) and the depth scaling factor (δy) which, together with the real diffusion time, define D_a.

7.2 The Parameter Optimisation

There are several ways of presenting a comparison between measured and computed modes of slab waveguides. Among these are the following:

Method (i). The measured values of n_e may be plotted against diffusion time and compared with a computed dispersion characteristic; this assumes that diffusion depth varies as the square-root of diffusion time. The W.K.B. method is used.

Method (ii). The method of Stewart,[44] which is to plot the W.K.B. integral against the effective index for both the measured modes and the computed profile. The W.K.B. dispersion equation is:

$$k_0 \int_0^{a_e} \sqrt{n^2(y) - n_e^2}\, dy = m\pi + \frac{\pi}{4} - \frac{\Phi_0}{2} \qquad (7.2.1)$$

where Φ_0 is given by 2.1.9.

The right-hand side of 7.2.1 is calculated using the measured n_e values and the left-hand side, using the computed trial profile; if all the parameters are correct, these curves should coincide. When $n_e = n_1$ the integral is zero-valued, therefore, the curve intersects the n_e axis at n_1.

With both of the above methods, the profile scaling is chosen to match the computed curve to the experimental curve at one suitable datum point. This is done automatically by the computer programs which were written to expedite the work; thus the value of D_a normally remains implicit. The computed and experimental curves, therefore, always intersect at the matching point. The slope and curvature may then be adjusted to match two other points by varying n_1 and α.

Method (ii) is only useful when multimode slab guides are available but allows a rapid assessment of the profile shape to be made. Method (i) also allows single-mode slab data to be utilised to give the curvature of the dispersion curve. Both methods of presenting the results are used.

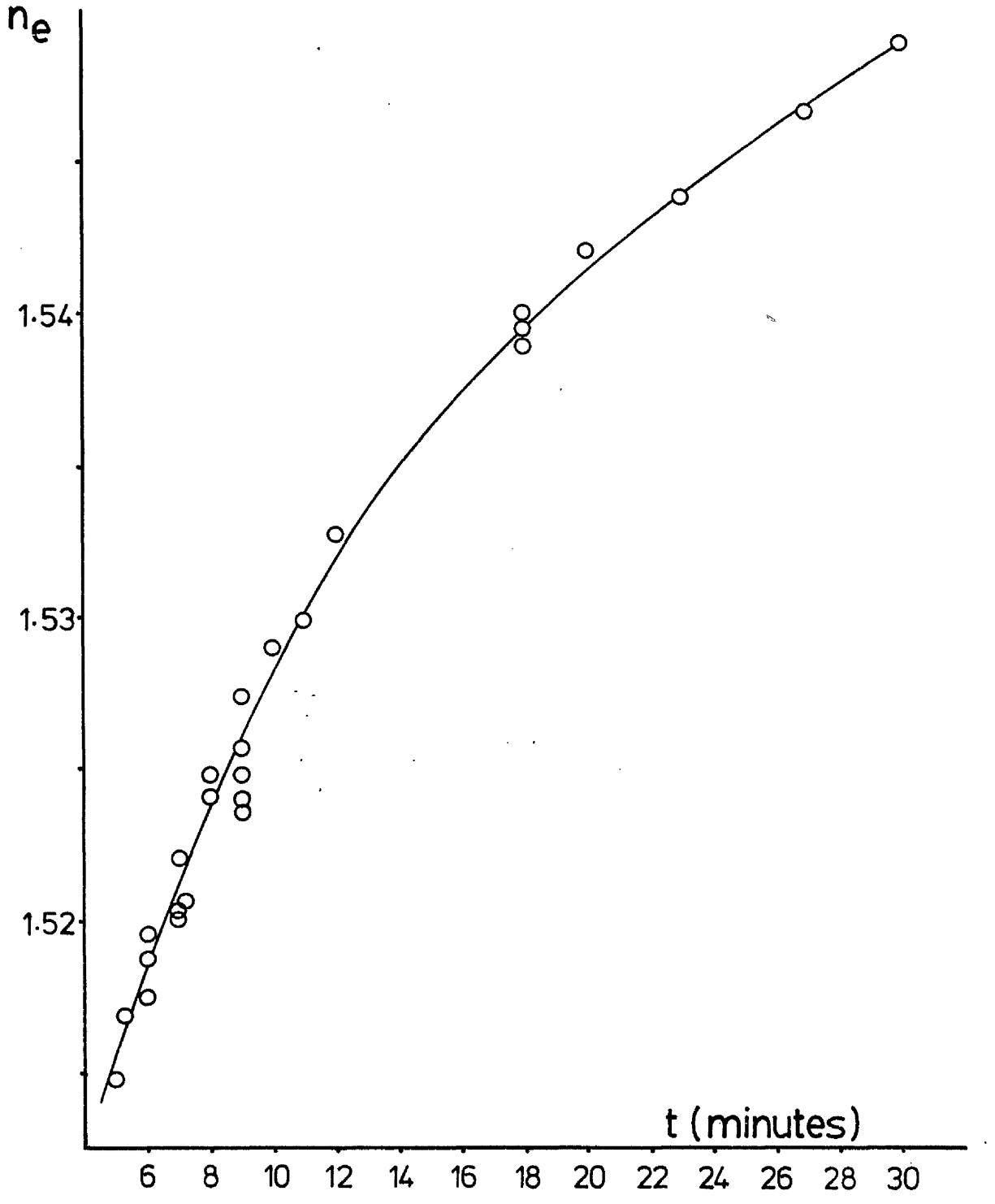


Fig. 7·1 Experimental graph of n_e vs. diffusion time[t] for slab, ion-exchanged waveguides

[T = 215° C]

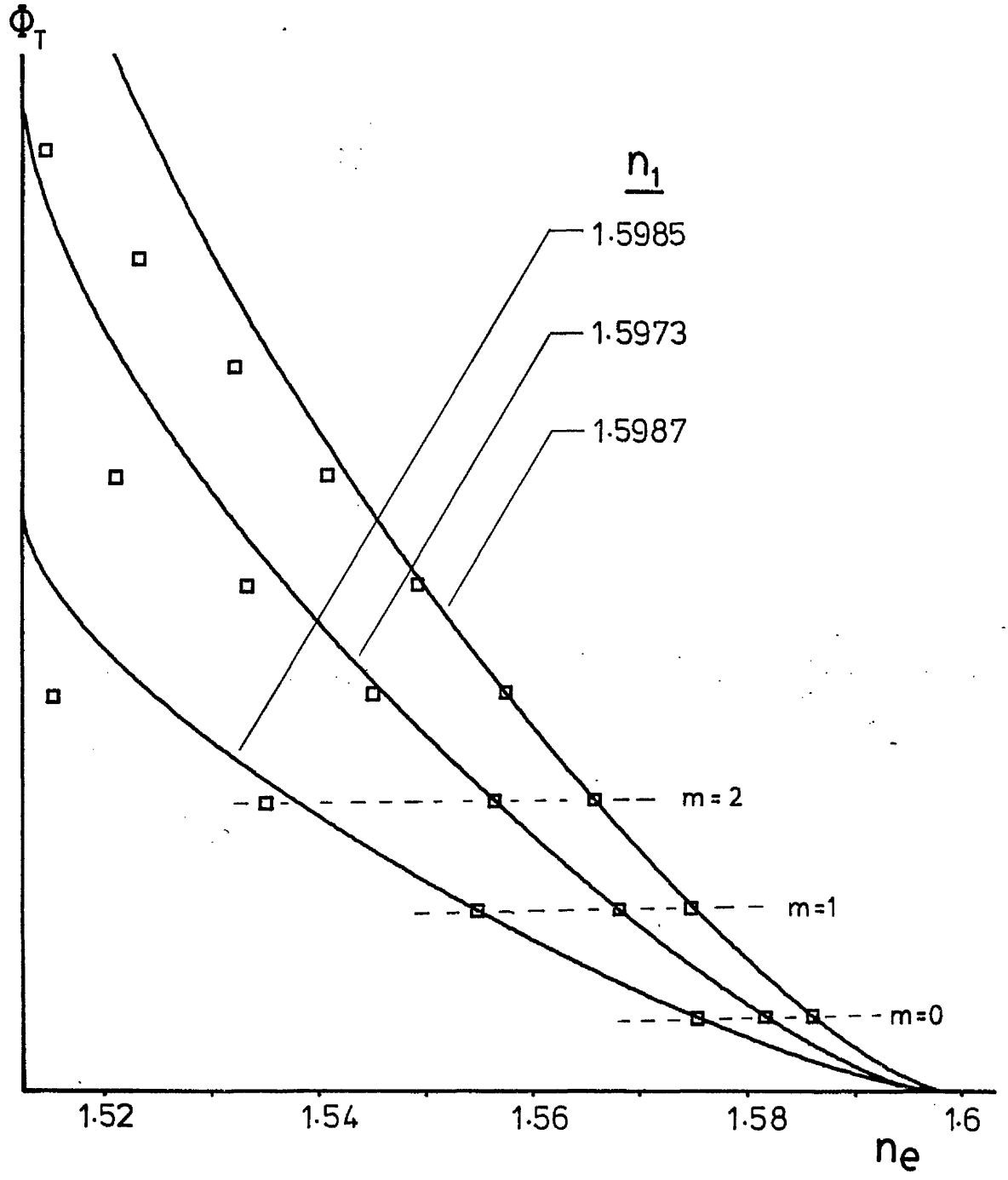


Fig. 7·2 Graph of W.K.B. integral [Φ_T] vs. n_e for :

measured n_e values □

computed diffusion profile \

$\alpha = 0.6$

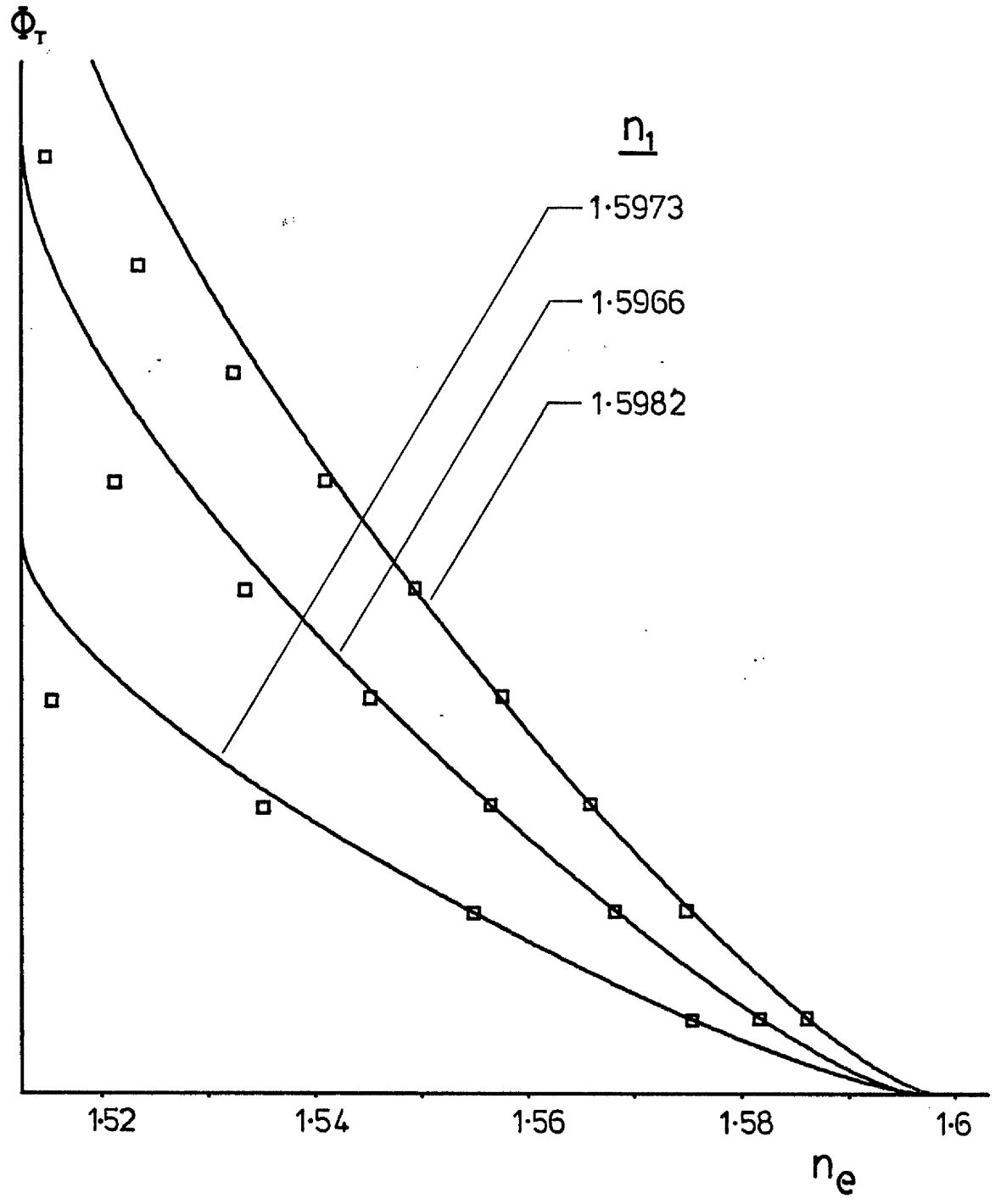


Fig. 7·3 – As Fig. 7·2

$\alpha = 0.7$

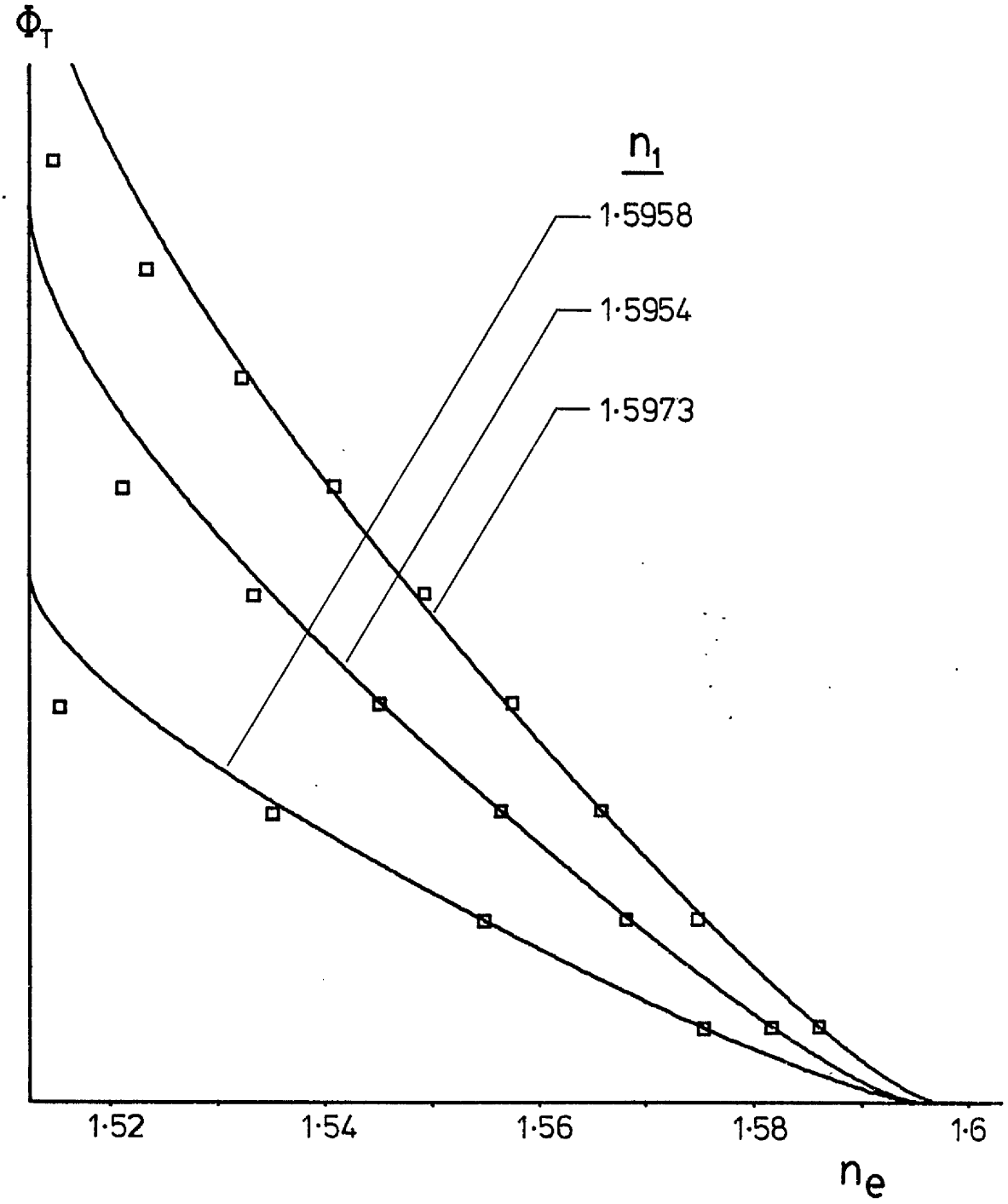


Fig. 7·4 — As Fig. 7·2

$\alpha = 0{\cdot}8$

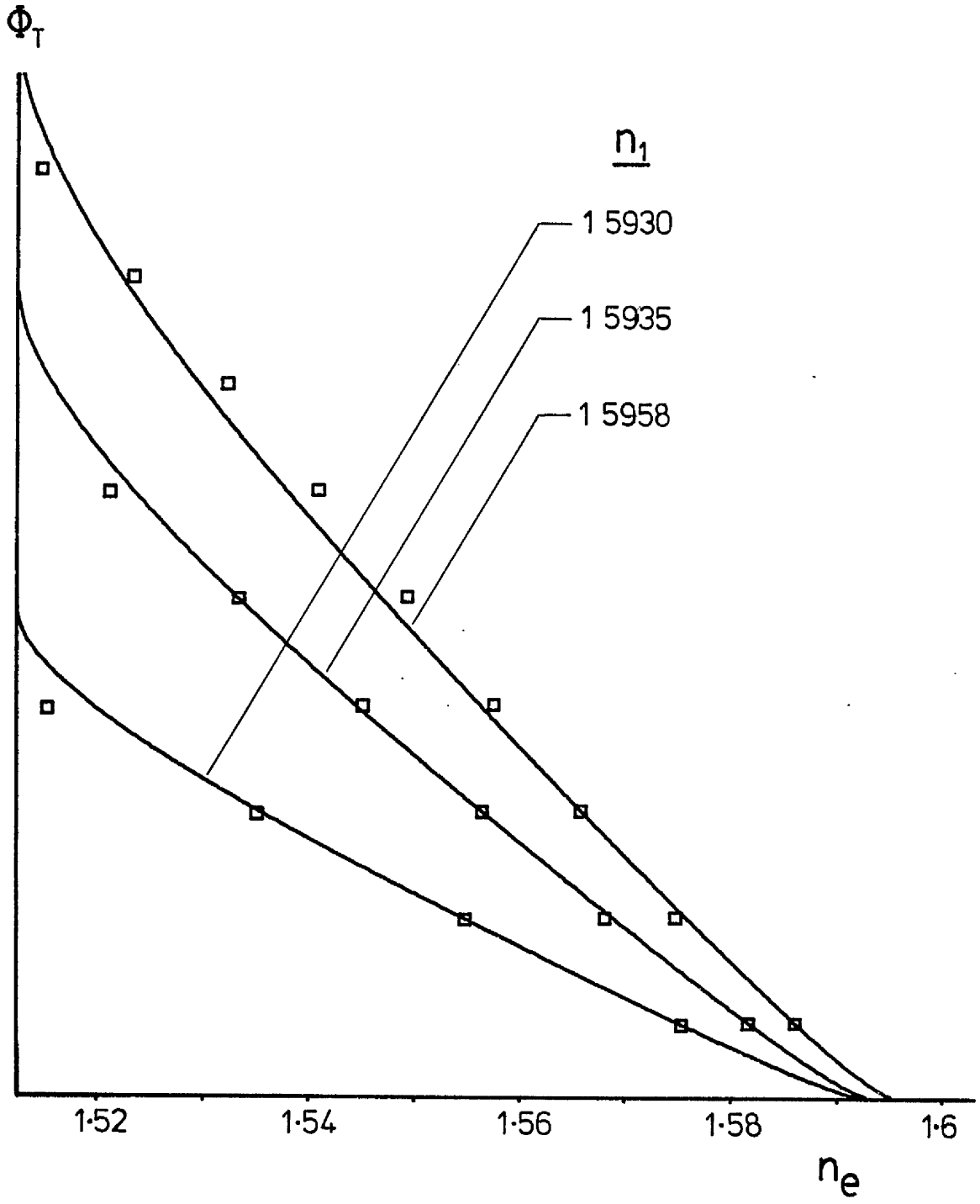


Fig. 7.5 — As Fig. 7.2

$\alpha = 0.9$

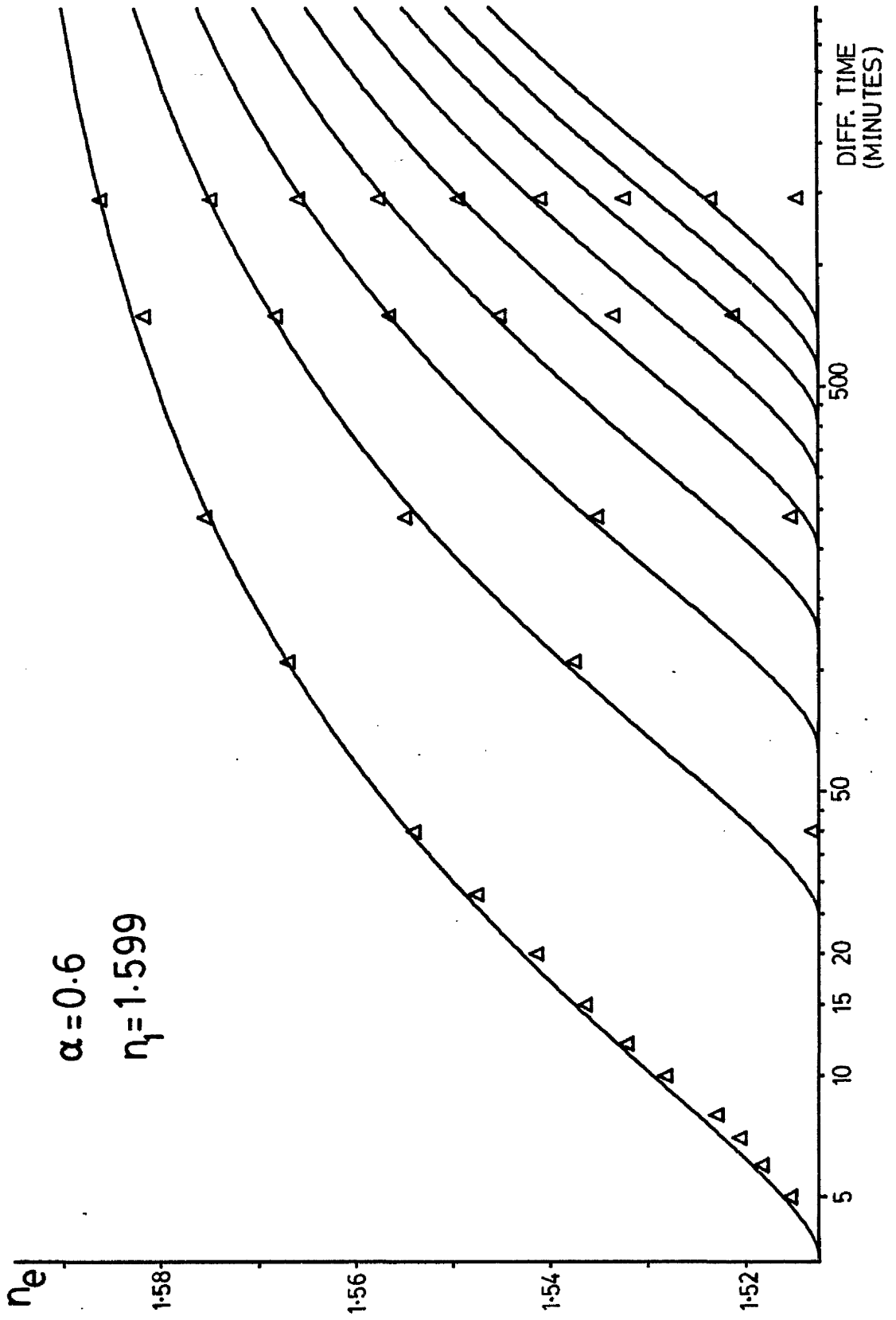


Fig. 7·6 n_e vs. diffusion time for experimental results [Δ], and for computed profile [W.K.B.]

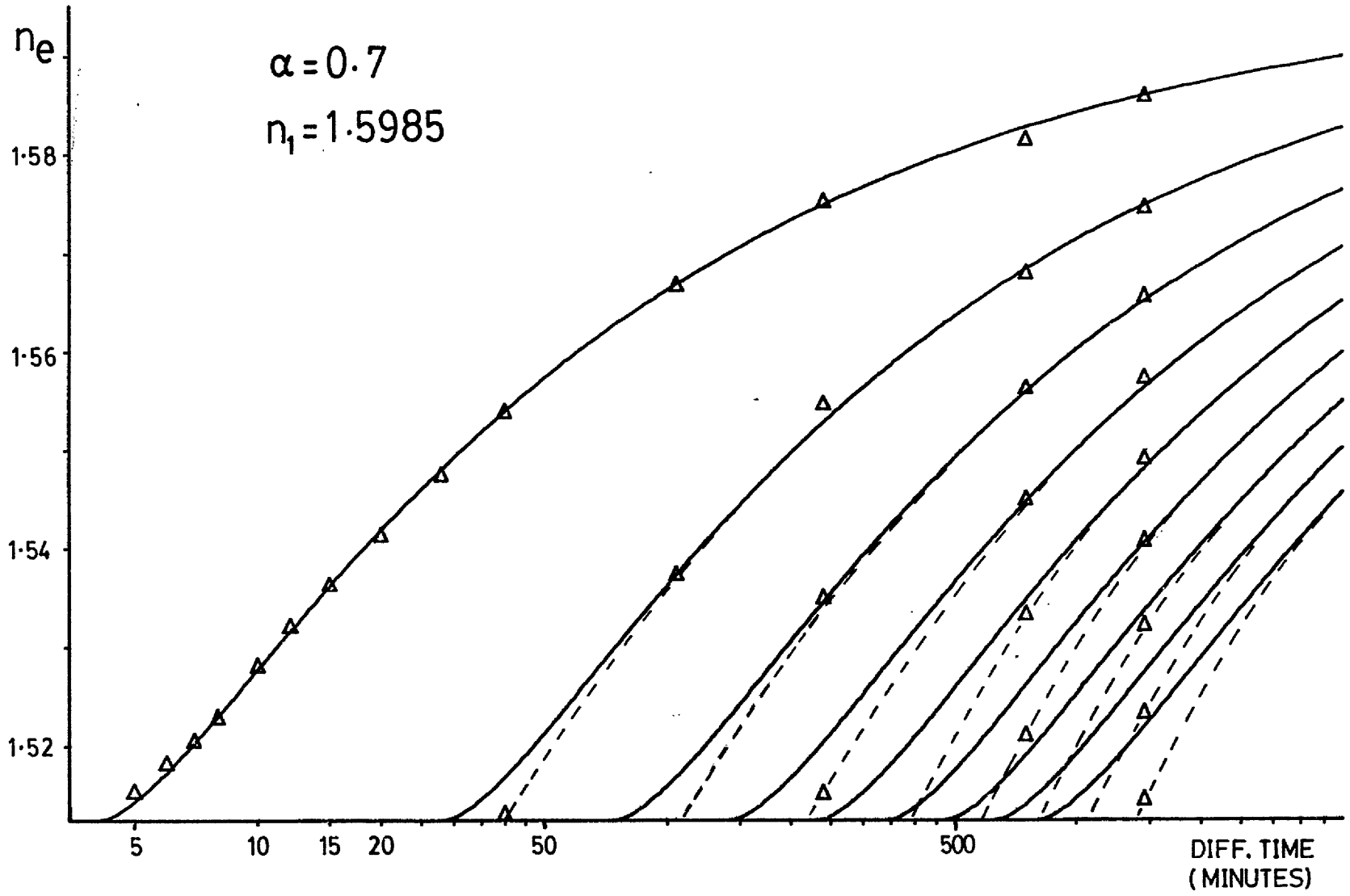


Fig.7.7 - As Fig.7.6

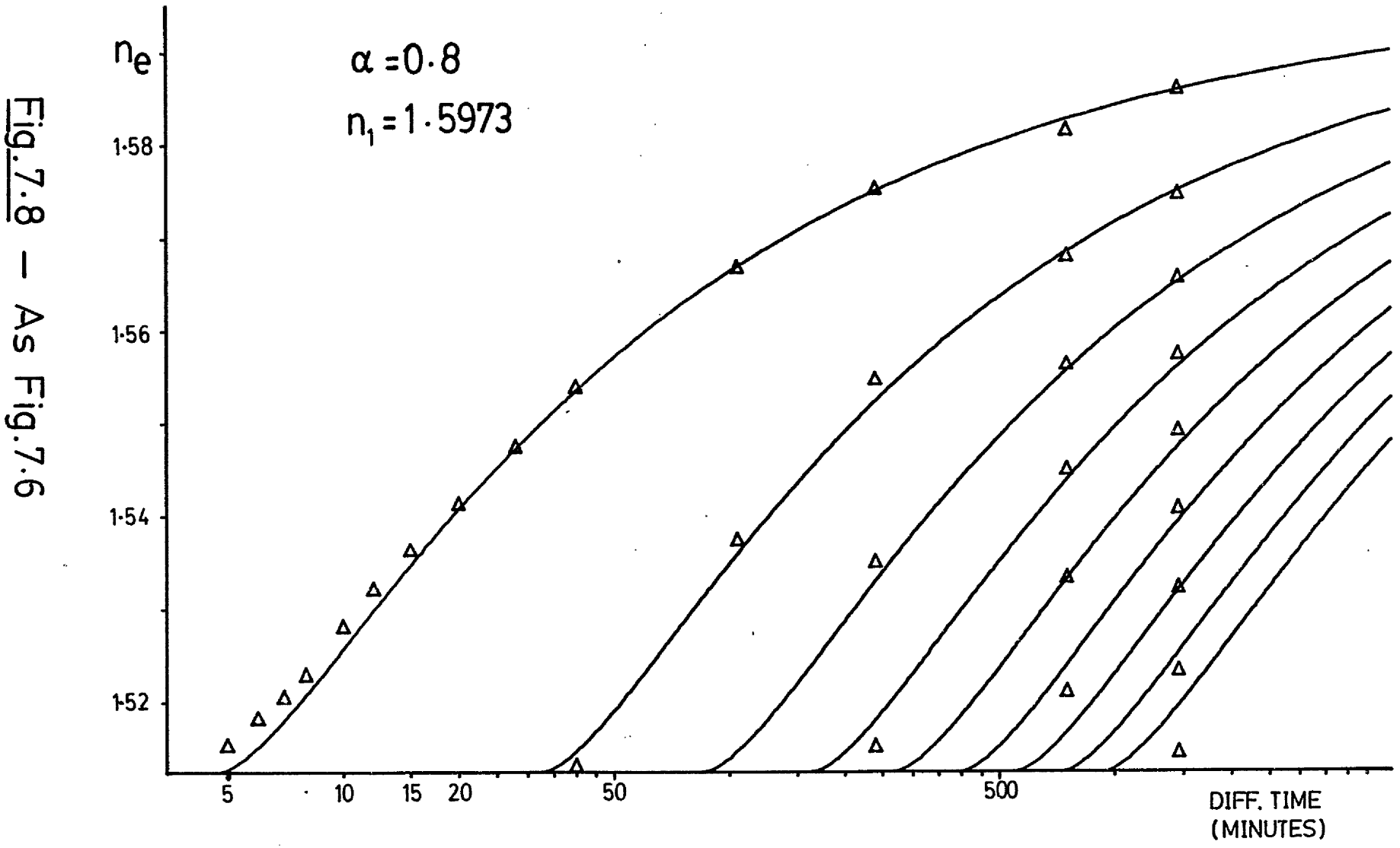


Fig.7.8 — As Fig.7.6

Results

Multimode slabs were diffused (at 215°C $\pm$ 0.2°C) for 1.75 hours (2 modes), 4 hours (4 modes), 12.5 hours (6 modes) and 24 hours, 21 minutes (9 modes).

The n_e values for single-mode slab guides are taken from the best-fitting curve through many experimental points whose scatter, below 12 minutes diffusion time, is considerable. The raw data for this is shown in Fig. 7.1. This scatter may, possibly, be attributed to slight temperature differences (the samples were not all diffused on the same day), to uncertain pre-heating of the slide, or to fluctuations in the composition of the surface layer of glass.

The three most highly moded slab guides are subjected to the analysis of method (ii) in Figs. 7.2-7.5 for α = 0.6, 0.7, 0.8 and 0.9. The results of method (i) are shown in Figs. 7.6-7.8 for α = 0.6, 0.7 and 0.8.

7.3 Analysis of Results

We consider first the data according to method (ii), as presented in Figs. 7.2-7.5; this reveals the basic properties of the profile shape. The following points are noted.

(i) An extrapolation of the experimental curve to the n_e-axis gives the physical value of n_1. We obtain this value by using the computed curve to perform the extrapolation, for which it must be matched to the three lowest order measured modes. As can be seen, this requires a low value of α, 0.6 or less, and gives a different value of n_1 for each guide. The standard deviation of these three n_1 values increases with α, confirming that the values become unrealistic for high α.

However, Figs. 7.6-7.9 suggest that the 6-moded guide really has a slightly lower surface index than the other two. The value of n_1 is about 1.599. Note that this value of α is not necessarily the true one.

(ii) It is possible to match theory to experiment almost exactly, for modes whose n_e is greater than about 1.55. Since the W.K.B. integral involves only that part of the profile for which $n(y) > n_e$, it follows that the upper halves [$n(y) > 1.55$] of the real profile and the computed profile, for $\alpha \simeq 0.6$, are the same shape. This is in accord with the findings of Stewart.[44]

The lower-half of the profile, on the other hand, cannot be matched by any value of α. The curve through the experimental points is nearly linear (for $n_e < 1.565$) and its curvature reverses sign; the computed curve, on the other hand, has a considerable unidirectional curvature.

The computed integral is too large; increasingly so as n_e falls, the implication being that the tail of the profile [$n(y) < 1.55$] is too extensive. The experimental profile, $n(y)$, seems to fall almost linearly towards the substrate value, n_s. This also is in accord with the findings of Stewart.[44] Large values of α (> 0.9), which deepen the upper half of the profile, dwarfing the tail by contrast, improve the average fit, as may be expected. However the shape of the curve is totally wrong, and unphysical values of n_1 are required to obtain a good match.

These observations are supplemented by a consideration of the data presented according to method (i) (Figs. 7.6-7.8).

In each case the fundamental-mode curves are matched at 105 minutes diffusion time; n_1 is then adjusted to match the fundamentals at 1461 minutes also. Since we are concerned as much with the time variation as with the mode distribution in this case, some compromise in the latter

must be accepted. There may be errors in the effective diffusion times, as the furnace temperature is prone to slight drift during long (e.g. 24 hour) diffusions. Thus, the chosen values of n_1 differ slightly from those of method (ii).

As can be seen, the graph for $\alpha = 0.7$ and $n_1 = 1.5985$ provides an almost exact fit to the experimental data for the fundamental mode, the curves for $\alpha = 0.6$ and $\alpha = 0.8$ showing too little, and too much curvature respectively. Fig. 7.7 ($\alpha = 0.7$) also seems to show the best average fit for the other two or three lowest order modes.

The question arises as to why the fundamental mode can be accurately matched right down to cut-off when the higher order modes cannot; the measured modes falling below the theoretical for small values of $n_e - n_s$. To answer this we consider Fig. 7.7 in which the divergent areas of the experimental curves are shown (approximately) dotted. When the mth experimental mode is at cut-off, the mth theoretical mode (for the same diffusion time) has an effective index, n_e^m (say). There is more than a suggestion that n_e^m increases as m (and hence, diffusion time) increases; especially between m = 1 and m = 3. Thus, it is possible that n_e^0 is close to n_s, and the fundamental mode can be matched almost down to cut-off. W.K.B. theory, considering two fixed profile shapes, cannot account for this behaviour and predicts a constant n_e^m. This raises the possibility that the profile shape is changing with time, and starts to lose its "tail" at perhaps 15-20 minutes diffusion time. Only the fundamental-mode, experimental dispersion curve shows any sign of the inflection point, that all the computed curves exhibit.

From this evidence, it is concluded that suitable computational parameters would be:

$$\alpha = 0.7$$

$$n_1 = 1.599$$

$$D_a = 1.02 \times 10^{-16}\ m^2/sec$$

D_a is calculated as described in section 5.3. This exact value will probably not be used in practice. Because of the large experimental scatter at short diffusion times, the measured slab n_e is used as the depth-defining parameter, rather than diffusion time.

We use the probable physical value of n_1, and compromise slightly in the value of α to improve the match further down the profile.

7.4 Post-Baked Slab Waveguides

Details of the experimental method for post-baking may be found in section 6.4.

The W.K.B. method and the variational computer program, SLAB, give slightly different results as the fundamental slab mode approaches cut-off (see Fig. 5.8). Therefore, because only one mode is present in most of the post-baked samples, and because the results are required to be compatible with the stripe guide model, the variational method is used in this section.

Fig. 7.9 presents computed curves of n_e vs. diffusion time (t) for the initial diffusion and two total post-bake times of t and 3t. The computational parameters are: $\alpha = 0.7$, $n_1 = 1.599$, $D_a = 1.02 \times 10^{-16} m^2/sec$ and the three profiles used are those plotted in Fig. 5.5. It is encouraging to note that the computed fundamental-mode curve of Fig. 7.9 is virtually identical to the best-fitting curve of Fig. 7.1.

Of the large number of devices whose slab measurements comprise Fig. 7.1, only a few were post-baked; the experimental results for these are also shown on Fig. 7.9.

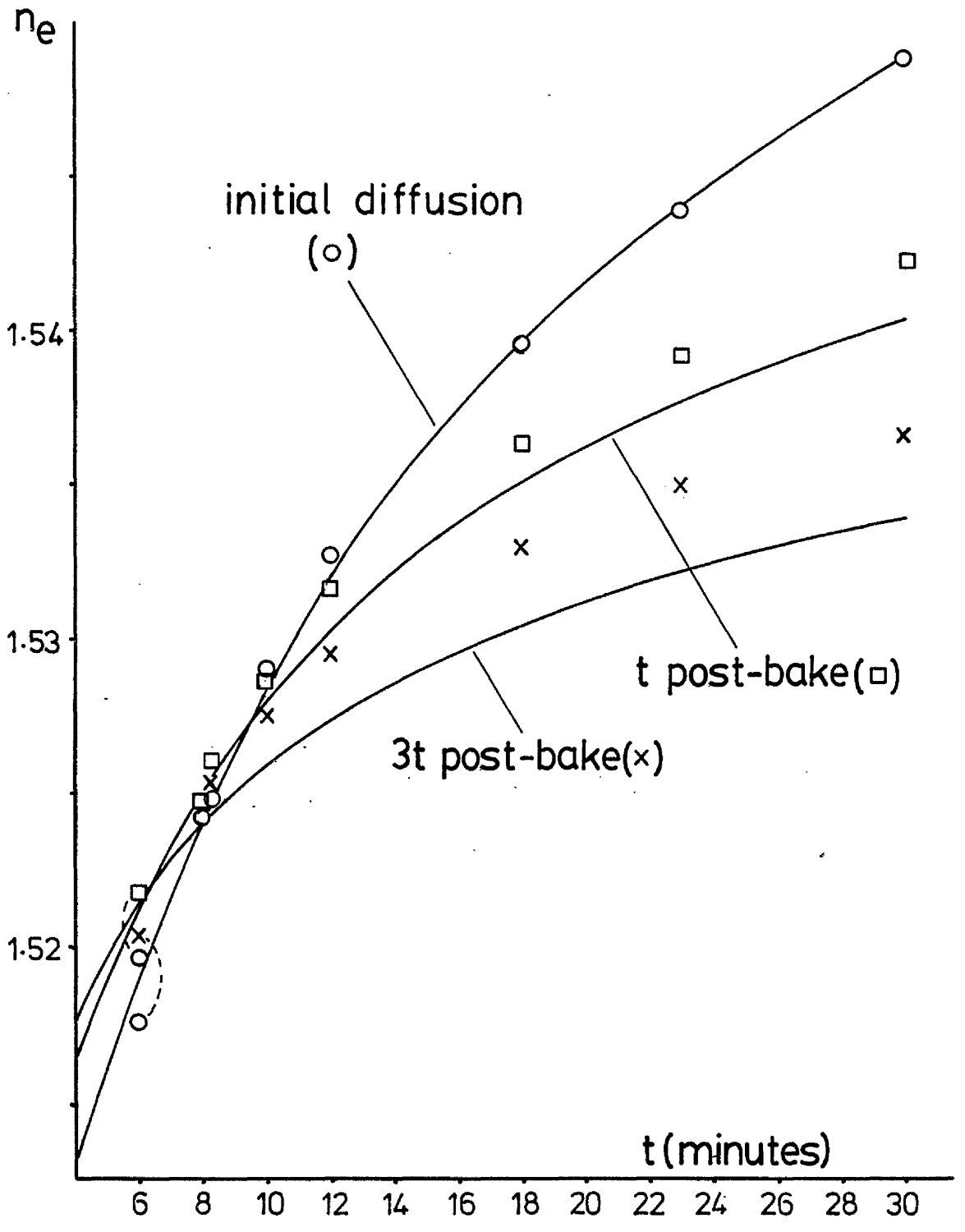


Fig. 7·9 Theoretical curves[variational solution] of n_e vs. initial diffusion time[t] for post-baked slab waveguides

Experimental results shown: ○ , □ , ×

Discussion

It can be seen that the experimental and computed results agree qualitatively, but that the quantitative agreement is limited. The two essential results of post-baking are that the surface index is reduced and the waveguide effective depth is increased. Modes near cut-off are more sensitive to the latter effect and n_e will increase at first. Well guided modes (large n_e) are more sensitive to the reduction of n_1, so n_e falls. Thus the mode separation decreases and modes may be gained. If the post-bake time is always equal to the initial diffusion time, then there will be a diffusion time (t) for which post-baking does not change n_e. This may be called the pivotal diffusion time, t_p. This, (at 9.5 minutes), is of interest for an entirely different reason. Coincidentally, a 2 micron wide stripe waveguide is only single-moded for diffusion times less than about 10 minutes.

The computed and experimental curves agree (approximately) about the value of t_p, but in general the computed values of n_e change much more rapidly, as post-baking proceeds, than do the experimental values. No changes in the computational parameters can entirely correct this, but a large increase in α is found to improve the fit somewhat. However, as discussed earlier, a large value of α requires an unrealistically small value of n_1 to obtain a fair match.

During post-baking, the peak concentration (normalised) soon falls to less than 0.6. Because of the form of the inter-diffusion coefficient, $\frac{D_a}{1 - \alpha u}$, this means that the post-bake diffusion is slower than the initial diffusion. Increasing α increases this rate differential, as required to improve the model, but spoils the model in other respects.

It was noted in section 6.4 that silver seems to be lost from the glass, appearing as a yellow surface coating, during post-baking. This cannot explain the discrepancy between the observed and the theoretical

behaviour as the error is in the wrong direction. (It may be added that waveguides which have not been post-baked tend to develop a similar yellow coating during storage. This may be due to atmospheric humidity and confirms that the surface, silver concentration is high; the profile is not buried.)

7.5 Results for Polished Slides

Several slides were ground and polished to an optical finish on one side in order to remove large surface defects. The effect of this on the waveguide losses will be considered in a later chapter; here the effect on the waveguide modes will be noted.

The four multimode slab waveguides discussed in sections 7.2 and 7.3 were made on the unpolished sides of such slides and are quite typical of untreated slides. The slab waveguides on each polished side, however, are found to have higher effective indices than the above, the difference being virtually constant for each slide but increasing slightly as diffusion time is reduced. This difference, Δn_e, lies between 0.0006 and 0.001.

When the data is compared with the theory according to method (ii) (see 7.2) it is found that the conclusions concerning α and the profile shape still apply, but that the surface refractive indices must all be increased by approximately the same amount, $\Delta n_1 = 0.0006$. Very short diffusions also show this effect quite strongly.

7.6 General Discussion and Conclusions

There are three general phenomena which require explanation:

1) The missing "tail" of the real diffusion profile.

2) The reduced diffusion rate during post-baking.

3) The increased surface refractive index when the slides are polished before diffusion.

Both (1) and (2) could be explained by a self diffusion coefficient, D_a, which varies with concentration; certainly the inter-diffusion coefficient is not of the correct form to match the experimental data. This is insensitive to α and u when u is small, so it is not possible to alter the tail region much by varying α. We require an inter-diffusion coefficient which falls steeply as u tends towards zero, rather than levelling off as does that derived in Chapter 4. This calls for an examination of one of the basic suppositions of Chapter 4: that D_a is constant.

Fick's diffusion laws assume a totally random ionic motion. If, however, there is some correlation, then an additional thermodynamic factor must be introduced. The inter-diffusion coefficient becomes:[50]

$$\frac{D_a}{1 - \alpha u} \frac{\partial \ln(A)}{\partial \ln(c_a)}$$

where A is the activity of the silver ion as a function of c_a (the silver ion concentration).

Doremus[50] assumes the form:

$$A = c_a^n \qquad \text{where n is constant}$$

with the result that D_a is simply multiplied by a constant (n), his results being consistent with a value of 1.1 for n. If the above function, $A(c_a)$, were of a slightly different form, this thermodynamic factor would

not be a constant. (A correlation giving the observed effect might be, for example, an inhibition of the silver ion mobility by a preponderance of sodium ions.)

Another possible cause of parameter changes is stress in the glass caused by ion-exchange itself. Doremus[40] notes that such stress can change the properties of the glass, including those important in ion-exchange.

Perhaps the most likely cause of the observed discrepancies is inhomogeneity of the glass. The results for polished slides show that optical working changes the surface refractive index, which indicates compositional change. It is almost certain that all glass surfaces, which have been exposed to aqueous solutions, differ in composition (and structure) from the bulk of the glass beneath. Such surfaces become "hydrated": silicon-oxygen bonds are broken and hydroxyl groups are formed; this can drastically alter ion exchange properties,[40] increasing diffusion rates, considerably. In addition, sodium is leached from the glass[41],[42] and hydrogen ions diffuse in to replace it. Such H^+ ion transport is known to occur in the glass electrodes of pH meters.[40]

We might, therefore, postulate the presence of a surface layer, on the cleaned glass, whose diffusion properties are graded; D_a falling from a high surface value to the bulk value. The effect of this depth dependence of D_a would be similar to the concentration dependence discussed earlier. However, for short diffusion times, $D_a(y)$ would vary slowly compared with $u(y)$ and the assumption of constant D_a would be more accurate; especially if $\frac{\partial D_a}{\partial y}$ were small near the surface. In general, the lower concentration values (being deeper) would encounter a reduced value of D_a, and the tail of the profile would be attenuated; only if the profile were deep enough to reach the homogeneous glass would the tail

reappear. An example of this is the experimental profile of Doremus,[50] measured by a radioactive tracing technique. In order to obtain sufficient accuracy, a very deep profile is required, and is about 8 times deeper than the 24 hour diffusion discussed here. The profile shape is given by 8 points (u against depth) with some scatter; there seems to be a tail to this profile, though it could still be smaller than that of the theoretical curve.

Thus, an inhomogeneous surface layer could conceivably account for some of the observed phenomena. Most of this discussion is, admittedly, speculative and the matter requires deeper investigation. It is clear from the literature that the mechanism of ion-exchange in glass is still very little understood.

Conclusions

The slab waveguide behaviour points to a value of n_1 of 1.599 and suggests that, although no value of α exactly models the diffusion profile, a value of 0.7 offers a suitable compromise. The discrepancy may be caused by laminar inhomogeneity of the glass surface, in which case the lateral characteristics of stripe waveguides may be only slightly affected by it.

Bearing in mind the uncertainty of the profile shape for short diffusion times, we accept the above value of α provisionally and proceed to a consideration of stripe waveguides.

CHAPTER 8

ION-EXCHANGED STRIPE WAVEGUIDES

This chapter contains experimental results, with theoretical comparisons, for stripe ion-exchanged waveguides. Both metallic aluminium and anodised aluminium diffusion masks are considered and a qualitative explanation of their behaviour is sought. The effect of post-baking stripe waveguides is also investigated.

In order to see how the real waveguides behave, as compared with the theory of Chapters 4 and 5, it is necessary to make a series of guides of different widths on the same substrate so that all will have the same slab effective-index. The waveguide pattern used to achieve this, shown in Fig. 8.1, has eight narrow stripes (2.5, 3.4, 4.2, 6, 8, 10, 15 and 20 μm widths), and a 120 μm wide quasi-slab. These all emerge from a common launching area. It was usual to create a true slab region beside the quasi-slab by means of additional rough masking during exposure.

The guides are made, and the widths and effective indices measured, as described in Chapter 6. The diffusion aperture widths are measured to an accuracy of $\pm$ 0.04 μm and n_e, to $\pm$ 0.0002, approximately. Both measurements are subject to fluctuations along the guide length, so efforts were made to ensure that the mode measurements were made at the same position as the width measurements, both being made the same distance along each stripe. The stripes were arranged to run across the slide width rather than down its length, so that any thermal gradients present during diffusion would affect each width equally.

8.1 Results for Aluminium Diffusion Masks

In the early stages of this work, the diffusion masks used were always of metallic aluminium; the anodised mask was developed later in response to the difficulties encountered. Thus it is logical to consider the aluminium diffusion mask first.

The measured effective indices of the first five lateral modes and for three diffusion times (7, 11 and 20 minutes) are plotted against the diffusion aperture width in Fig. 8.2. The corresponding computed curves (DIF and WAVE ; $\alpha = 0.7$) may be found in Fig. 8.3 for the same three slab n_e values. (The experimental points in Fig. 8.3 are for the best anodised masks; they may be ignored for the present, but will be considered later.) Concentration contour maps of the three 2 μm wide guides are shown in Fig. 8.4.

While these experimental results are slightly closer to our theory than are those of Gallagher,[55] the disagreement is still sufficiently violent to be clearly seen even without a superimposition of the graphs. The nature of the discrepancy may be summarised as follows:

(i) At any width, the measured n_e is smaller than the computed value. This difference increases as the width diminishes, resulting in an experimental dispersion curve whose slope is always greater than that of the computed curve.

(ii) Despite this, the mode separations ($n_{e_0} - n_{e_2}$ etc.) for each width are only slightly larger than those of the theoretical model. The fit, in this respect, is improved by using the $u = 0$ boundary condition [see Fig. 5.5 (b)] which makes the guides of the model a little narrower. The mode separations for various conditions are plotted in Fig. 8.7.

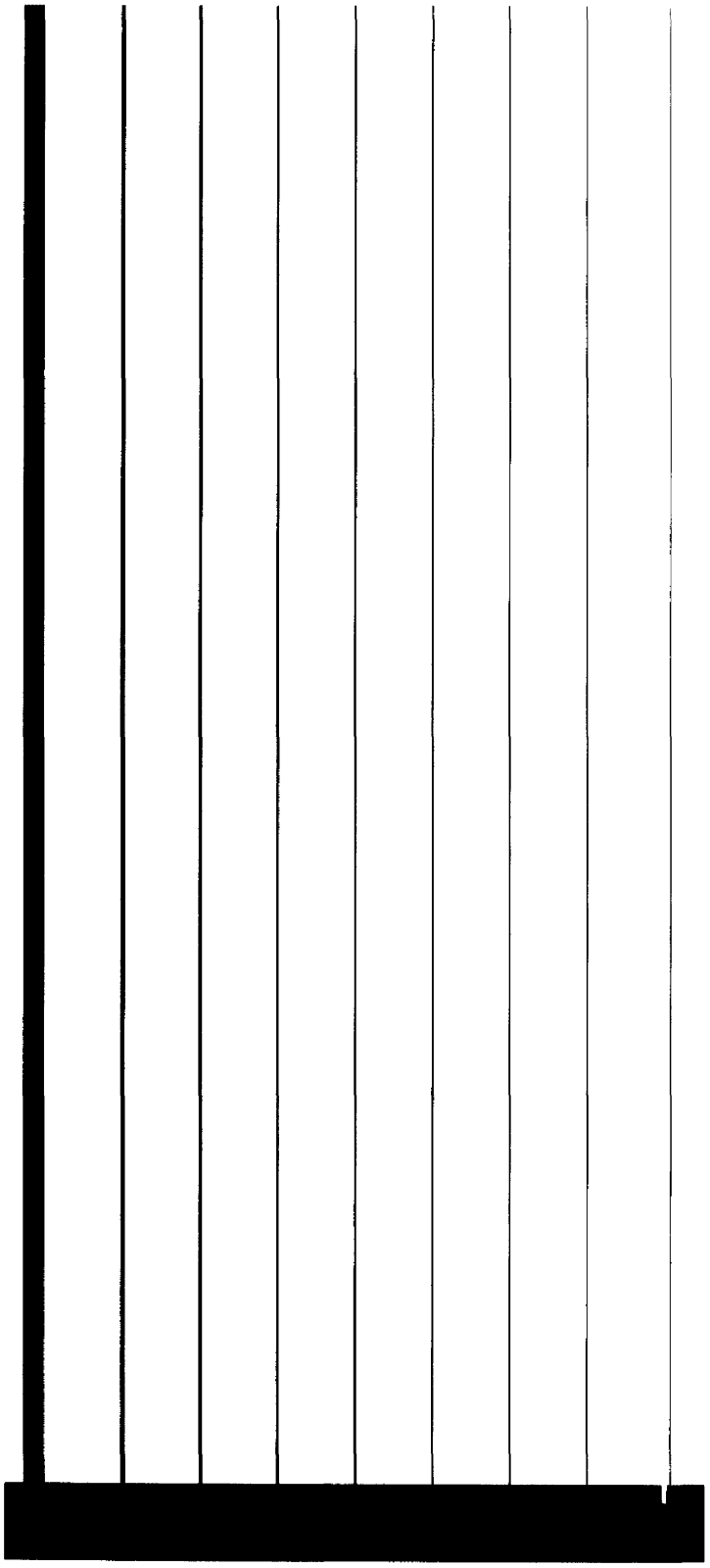


Fig. 8.1 The stripe waveguide shadow-mask

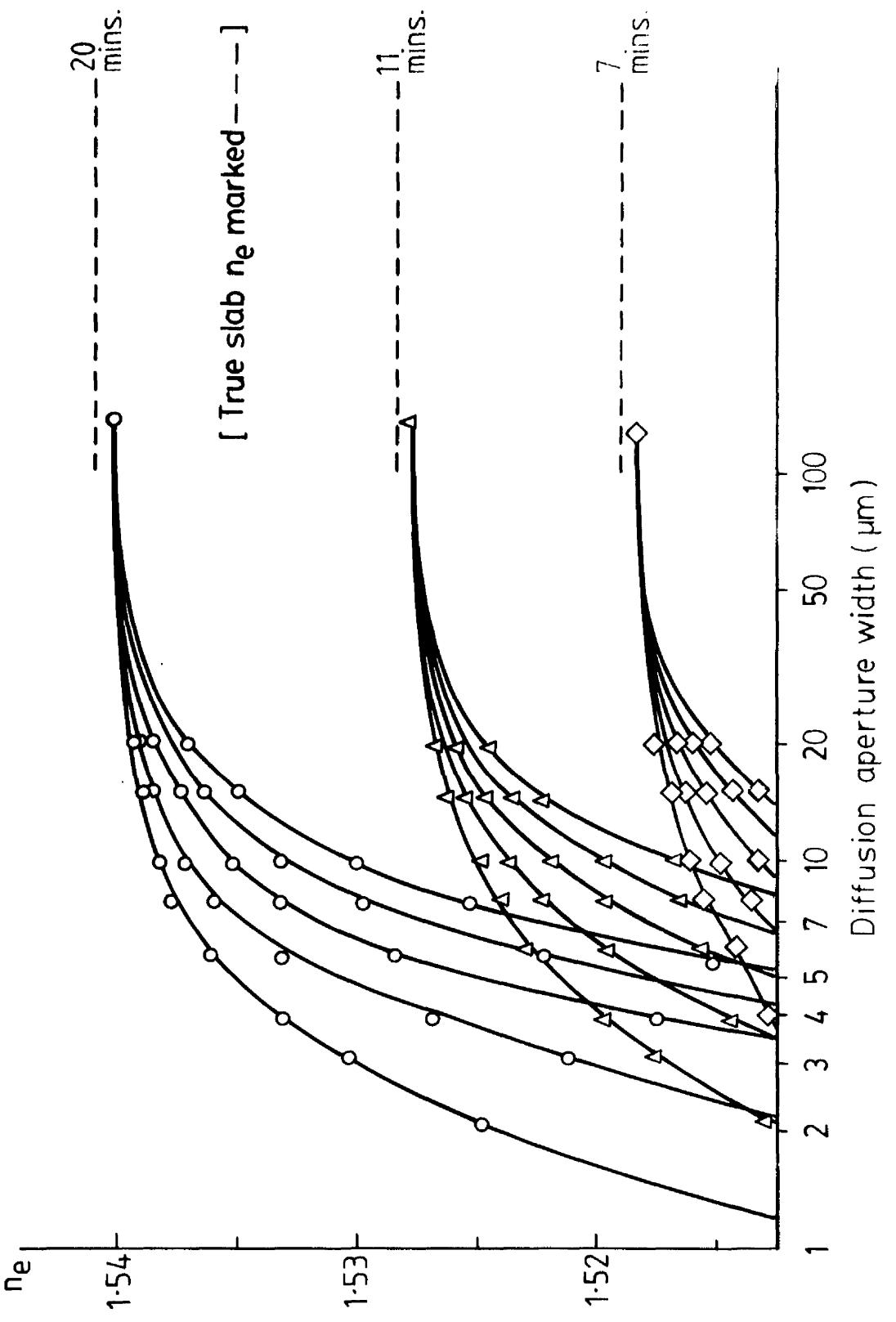


Fig. 8·2 Experimental curves of n_e vs. width for stripe ion-exchanged waveguides, made using ALUMINIUM diffusion masks.

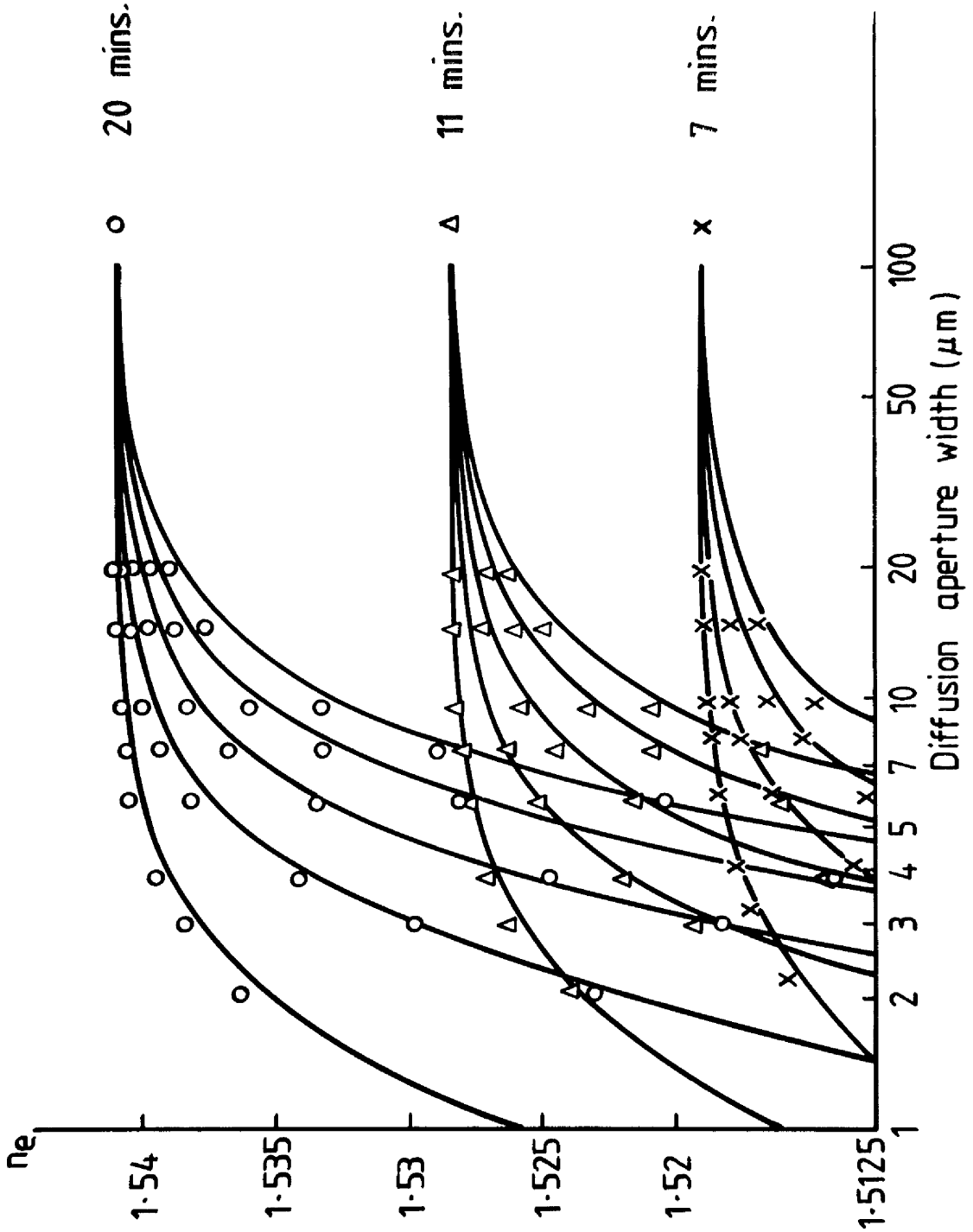


Fig. 8·3 Theoretical curves of n_e vs. width for stripe, ion-exchanged waveguides. The experimental points were obtained using the IMPROVED ANODISED diffusion mask.

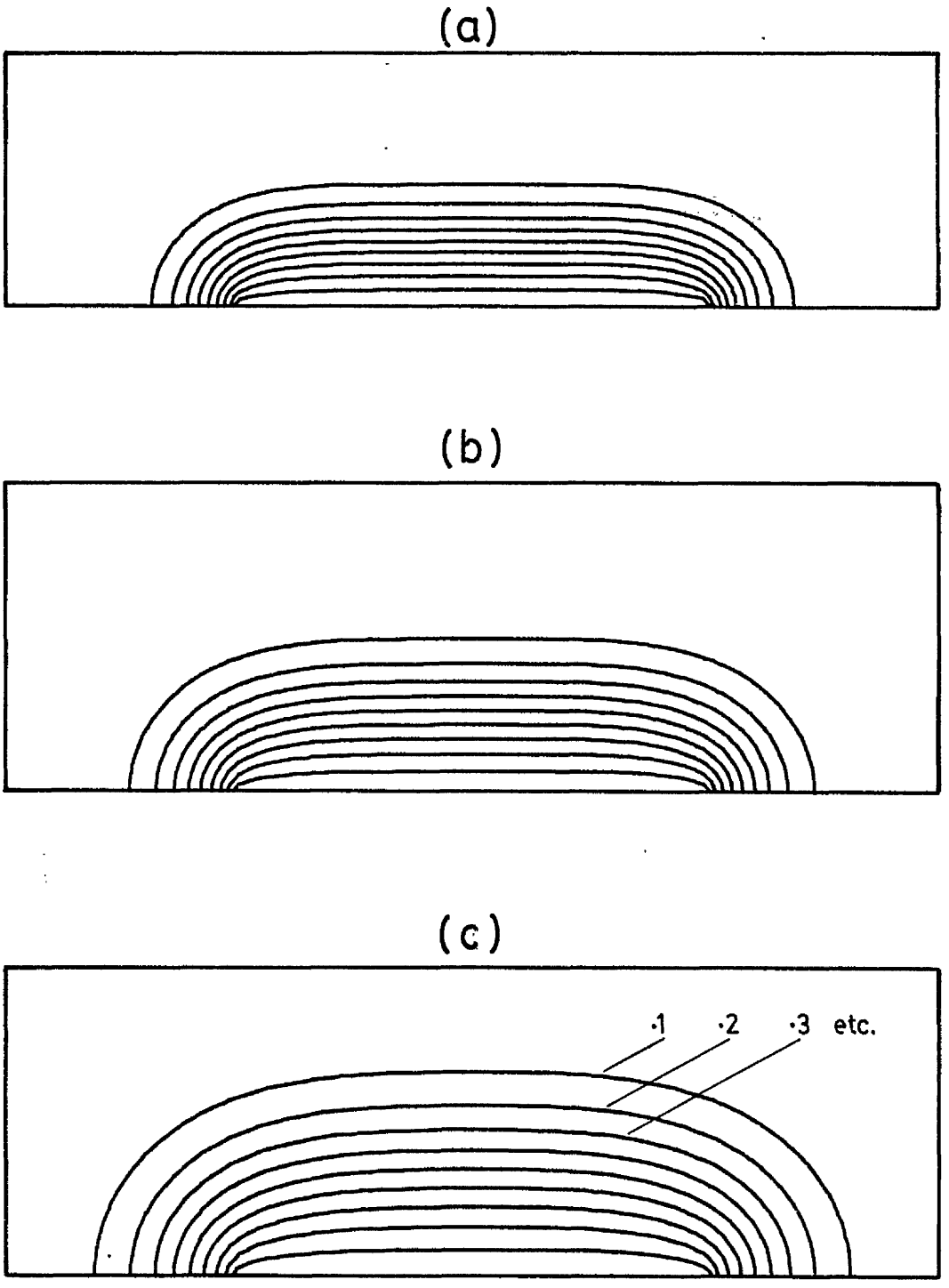


Fig. 8·4 Computed concntration contours
[u=0·1, 0·2 .... 0·9]
aperture: 2 μm
diffusion times: (a) 7, (b) 11, (c) 20
minutes

(iii) At a width of 20 μm, the computed fundamental dispersion curve has virtually reached its limit (the true slab value of n_e). The measured n_e, however, is less than this limiting value even at 120 μm width (the quasi-slab).

Adjusting the Model

There is no parameter which may be varied to match the computed curves to the experimental. Varying D_a, α or n_1 (the surface index) serves only to move the entire curve-set up or down the n_e-axis, the slopes and mode-separations remaining substantially the same. Thus, if α or n_1 is changed and D_a (i.e. the scaling) then adjusted to return the slab n_e to its original value, it is found that the lateral modes also return (almost) to their original values.

It is concluded that the mode separations are almost purely a function of width, and that the actual value of n_e at any given width is a function only of the effective diffusion depth and the surface refractive index. Thus, it may be concluded, from (ii) above that, though the measured widths are substantially accurate, the guide effective widths are a little narrower than predicted. From (i) and (iii) it may be concluded that either the diffusion depth or the surface refractive index (or both) diminishes as width is reduced. This effect will be called "width-dependent retardation".

Since D_a and α cannot be width-dependent, there are only two possibilities: either there is a width-dependent, external electric field retarding the diffusion,[46] or the surface silver concentration (in effect, the source concentration) is a function of aperture width. In the latter case both surface refractive index and diffusion depth are reduced by the same mechanism and the effect can be checked using DIF.

Adopting the u = 0 boundary condition to reduce the effective guide width, it is found that the computed modes may be matched to the measured modes almost exactly using the following (sample) values of u_0 (the normalised surface concentration):

Aperture Width	u_0
2.95 μm	0.89
4.35 μm	0.925
5.98 μm	0.95

Such empirical fitting is very crude and reveals little about the actual silver-ion distribution in the guide; however it confirms that the effect is indeed a width-dependence of the mean diffusion depth.

The above two possibilities may not be distinct, since such a reduction of u_0 may be electrical in origin. A major obstacle to any non-electrical explanation is the exceedingly long range of the influence. This is clearly associated with the diffusion aperture edges (growing stronger in the centre as the edges draw closer together) yet extends at least 60 μm from the edge to retard diffusion into the centre of the quasi-slab.

In order to check the effect of the mask conductivity, insulating diffusion masks were required. The idea of anodising the aluminium mask originiated with the anodisation of the high-purity, aluminium, silver nitrate bath (necessary to prevent its corrosion by the melt). The procedure is outlined in section 6.2.

8.2 Results for the Anodised Diffusion Mask

1) The Early Anodised Diffusion Masks

Fig. 8.5 presents typical experimental results for early anodised masks, showing dispersion curves for two diffusion times (7 and 20 minutes) for the five lowest order modes only. Later results, using improved fabrication techniques, approximate to the computed curves quite closely (see Fig. 8.3). However these early results are of great interest because, in almost every respect, the discrepancy between the measured and computed characteristics is equal to that found with the metallic mask, but in the opposite sense. To summarise:

(i) For any given width, the measured n_e is greater than the theoretical value, the difference increasing as width diminishes.

(ii) The mode separations are smaller than those of the computed curves (see Fig. 8.7), but the discrepancy is greater than for the metallic mask. The effective width is greater than predicted.

(iii) At 120 μm width, the measured n_e is greater than the theoretical, and shows no tendency to fall further as the width increases.

This characteristic is not exactly repeatable; however, frequently the theoretical dispersion curve is very nearly the average of the metallic and the anodised cases. When only half of the length of the guides is anodised, the transition between the two extreme characteristics is very abrupt.

Clearly, a narrow stripe guide cannot have a greater effective index than a slab guide unless it is considerably deeper. Thus, using the same reasoning as in section 8.1, we postulate a "width-dependent enhancement" effect in which the field of influence, which retards diffusion if the aperture is defined by a metal mask, is almost exactly reversed. If this influence is associated with the mask edges, it follows

that in the wider guides the edges will be more deeply diffused than the centre. This expectation is confirmed by the input m-lines, photographs of which are presented in Fig. 8.8 (fundamental modes only).

It is clearly seen that as the guides become wider the fundamental m-line increasingly resembles a single-slit diffraction pattern (cf. Fig. 6.12). The field distribution of the m-line is given, approximately, by the Fourier Transform of that in the waveguide. Thus, the usual single lobed m-line (see Fig. 6.8) indicates an approximately gaussian lateral field distribution. (This is confirmed by WAVE. See Fig. 5.13.) The m-lines of Fig. 8.8 indicate a field distribution which becomes more rectangular as the guides become wider, possibly even developing a shallow minimum in the centre. Comparison with the guiding structure and symmetric-mode field of a directional coupler confirms that the edges must be deeper than the centre.

This also explains the behaviour of the limit as the width tends to infinity. The energy propagates preferentially in the deeper edge regions irrespective of the width; the true slab n_e can be obtained only if the coupling spot is far from the edge.

2) The Improved Anodised Diffusion Mask

Although the first anodised masks produced a very interesting dispersion characteristic, the resulting waveguides were not of very high quality, having somewhat rough edges; also in places the silver nitrate had leaked through the mask. This was corrected by means of the more thorough cleaning procedure outlined in section 6.1, ensuring improved aluminium to glass adhesion. Several other changes were also made at this time, including a substantial reduction of the anodisation temperature, and a greater attention to cleanliness in general.

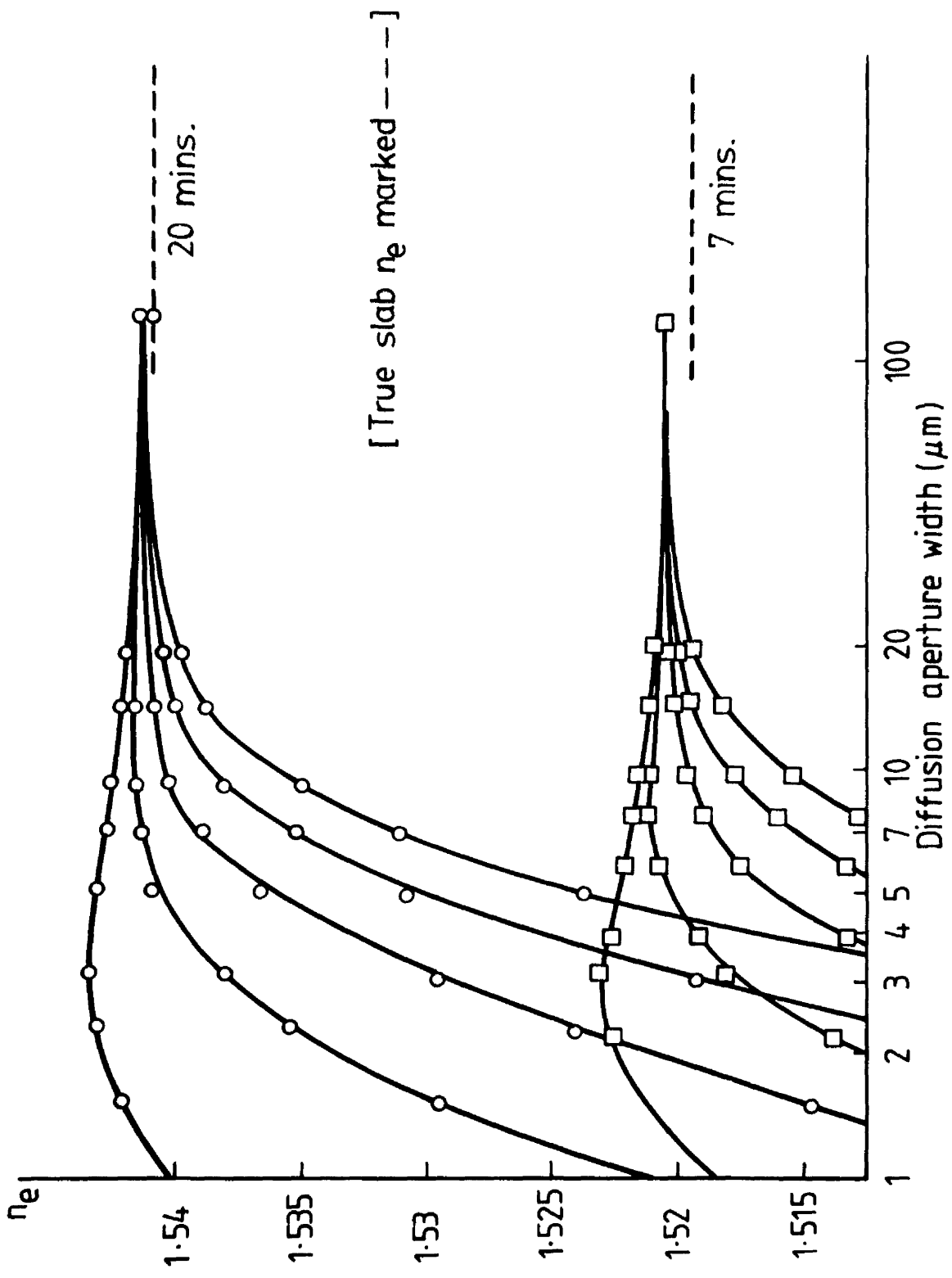


Fig.8·5 Experimental curves of n_e vs. width for stripe, ion-exchanged waveguides, made using (early) fully ANODISED diffusion masks.

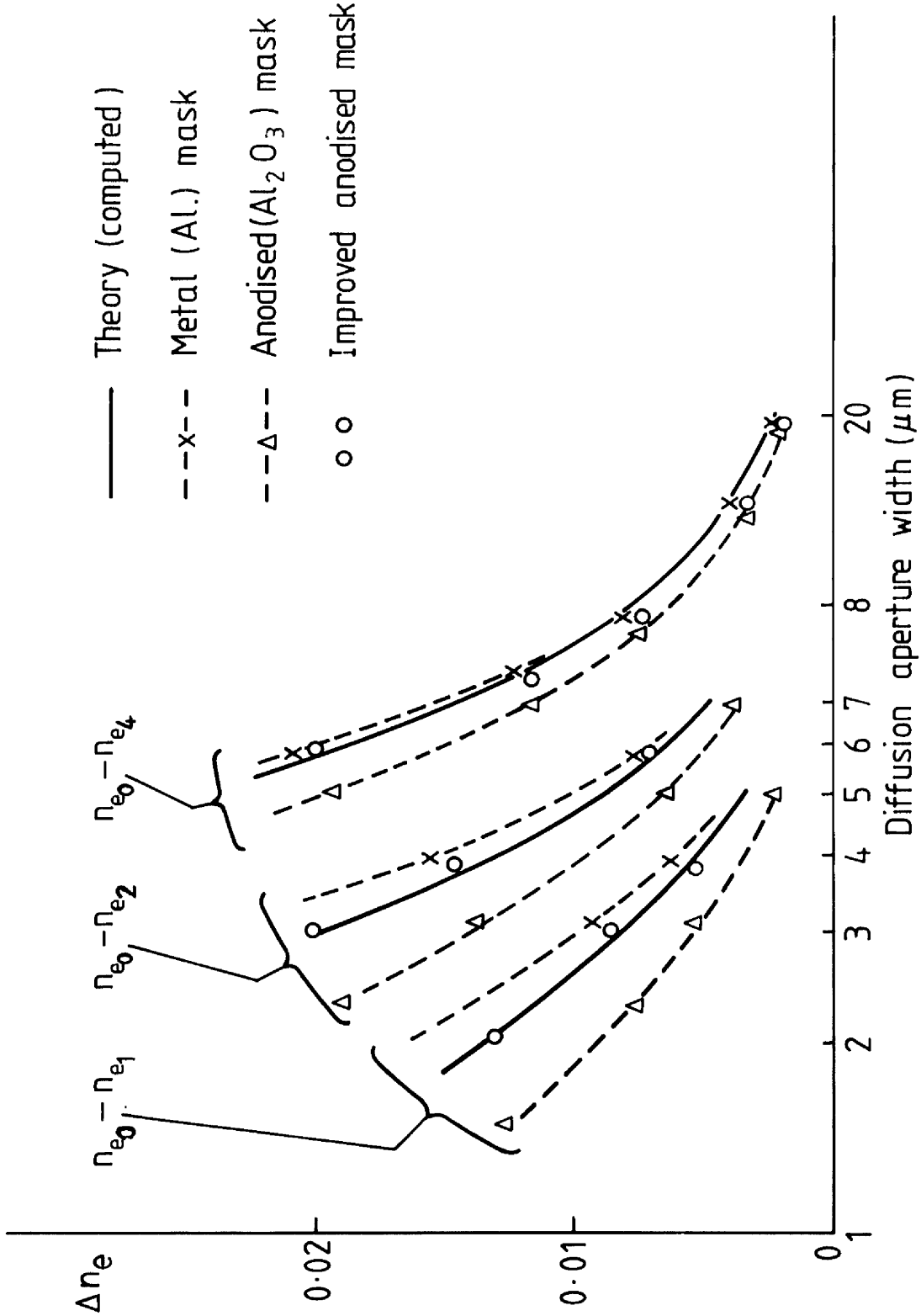


Fig. 8·7 Mode separations (Δn_e) for various diffusion mask types. 20 minutes diffusion.

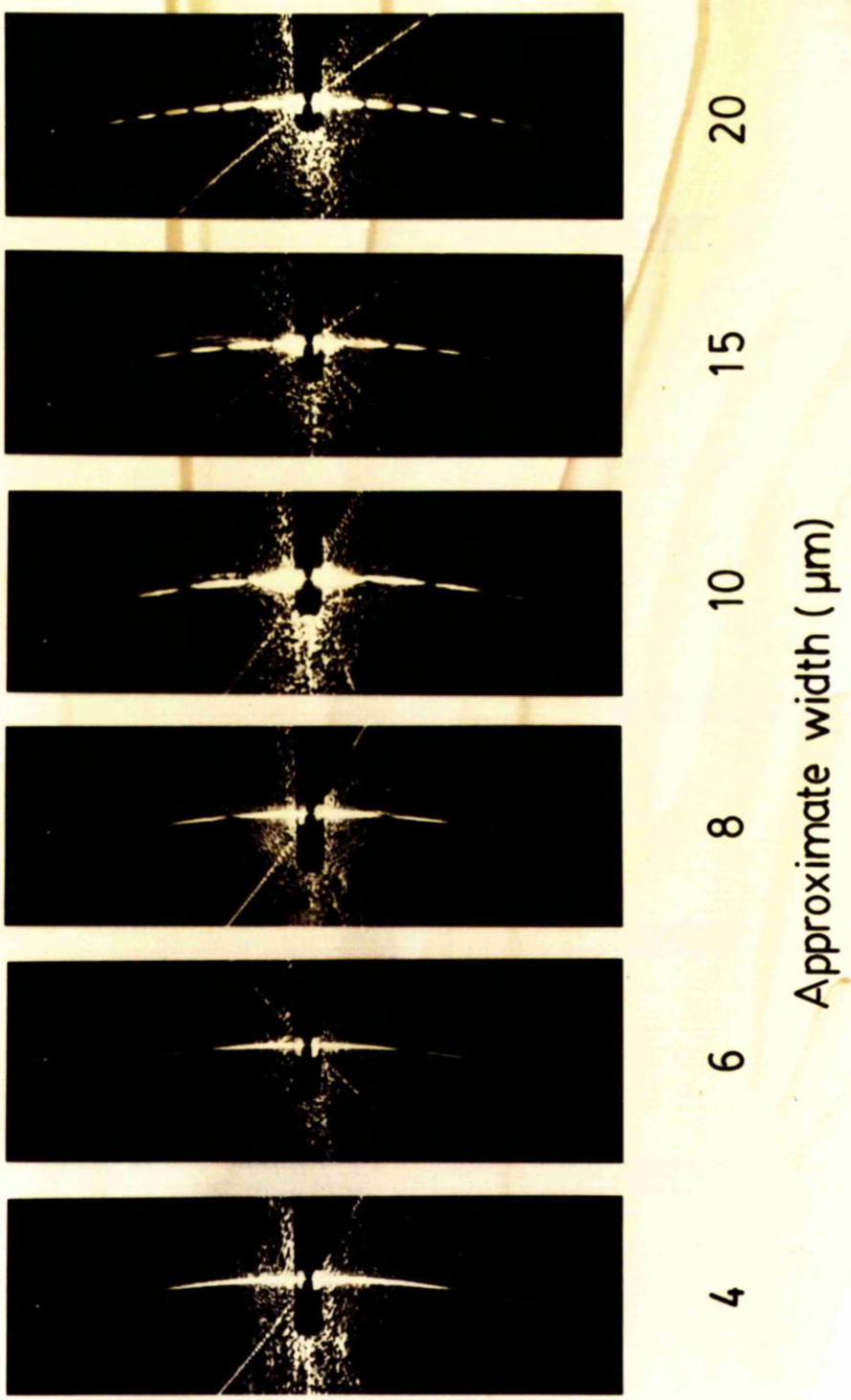


Fig. 8.8. Input m-lines (fundamental mode only) of stripe, ion-exchanged waveguides (widths as stated) made using an (early) anodised diffusion mask.

The result of this was not only a greatly improved masking quality, but also a change in the dispersion characteristic. The measured modes, though still slightly higher than the theory, now fitted the computed modes very closely. Fig. 8.3 shows such experimental points superimposed upon the computed curves.

The improved anodised characteristic proved to be quite repeatable; so repeatable that the early anodised characteristic could no longer be produced. Various changes in the fabrication procedure were tried (short of deliberately reducing the cleanliness) but the factor which determines the width dependent enhancement could not be isolated.

Some recent results have suggested that cleanliness alone is not the critical factor as was thought, but that the condition of the oxide film may depend upon how it is rinsed after anodisation. This ostensibly trivial operation was not recorded or controlled.

It was considered that the improved anodised mask matched the computer model sufficiently closely to enable both to be used with confidence. Thus, work on waveguide devices was undertaken and is discussed in Chapter 9. In the remainder of this chapter other diffusion-mask effects will be considered and an explanation of these phenomena sought.

8.3 Other Diffusion-Mask Effects

There are several other important observations to be noted.

1) Partial Anodisation

Two anodisation trials were performed, in which several sets of stripes were partially anodised by varying degrees. A typical graph of anodisation current vs. time is given in Chapter 6 (Fig. 6.3). The point 'A' marks the onset of catastrophic conductivity loss as the current falls

finally to zero. At A the anode current is I_0.

In both trials the measured modes were found to move up slowly towards the theory as the anodisation end-point was advanced towards A. However, as the end-point passed A, the measured modes rose rapidly towards their fully anodised positions - either "early anodised" (in the first trial) or improved (in the second trial).

Within limits, the position of the measured modes, relative to their final fully anodised positions seems to be directly related to the end-point current. This is itself (after the initial current oscillations) determined by the residual sheet conductance of the mask.

2) Metal Thickness

As may be expected from the above, the thickness of a metal diffusion mask affects the measured modes slightly. This effect has been observed, but not investigated thoroughly.

3) Association

When the mask is of aluminium, it is found that the proximity, or otherwise, of other unmasked, exchanging regions affects the measured modes of a waveguide. Two examples of this have been noticed:

a) If the mode measurements are made close to the open launching area of the waveguide pattern (see Fig. 8.1) the n_e values are elevated towards the computed values.

b) The difference between the true slab n_e and the quasi-slab n_e depends upon the width of the metal strip which separates the two exchanging regions. This effective-index difference tends to vanish if the partition is much less than about 1 mm wide.

4) Free Silver Reduction

This is the most important of the effects noted in this section, since it bears upon the usefulness of stripe waveguides made using an aluminium diffusion mask.

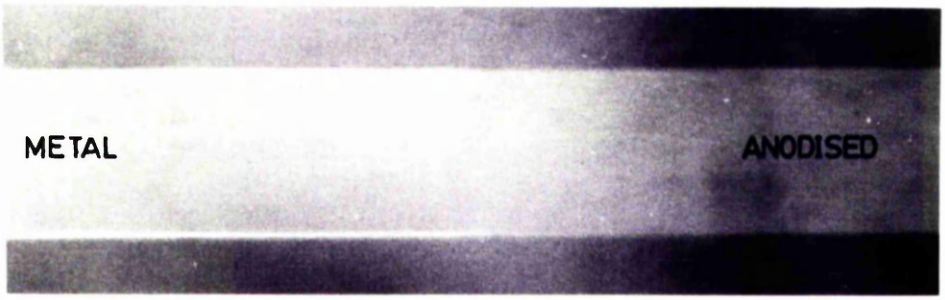
When the surface of such a waveguide is examined by means of a powerful optical microscope using reflected light, a cream coloured line can be seen running down each edge. This becomes broader and darker (eventually becoming brown) as diffusion time increases, and occurs only where the aluminium film and the diffused region have overlapped. This discolouration is assumed to consist of free, metallic silver which has been reduced by some action of the aluminium mask. It increases the opacity of the edge regions when transmitted light is used, but cannot be seen under transmission phase-contrast (see Fig. 6.5); it thus escaped detection for some time. It appears to be more than a surface deposit of silver, since nitric acid has no effect upon it.

Fig. 8.9 presents photographs (at 1000 times magnification) showing this reduced silver on the waveguides of an early ring-resonator pattern. The waveguides are 5 microns wide, the ring diameters are 200, 100, 50 and 25 microns, and the diffusion time was 50 minutes.

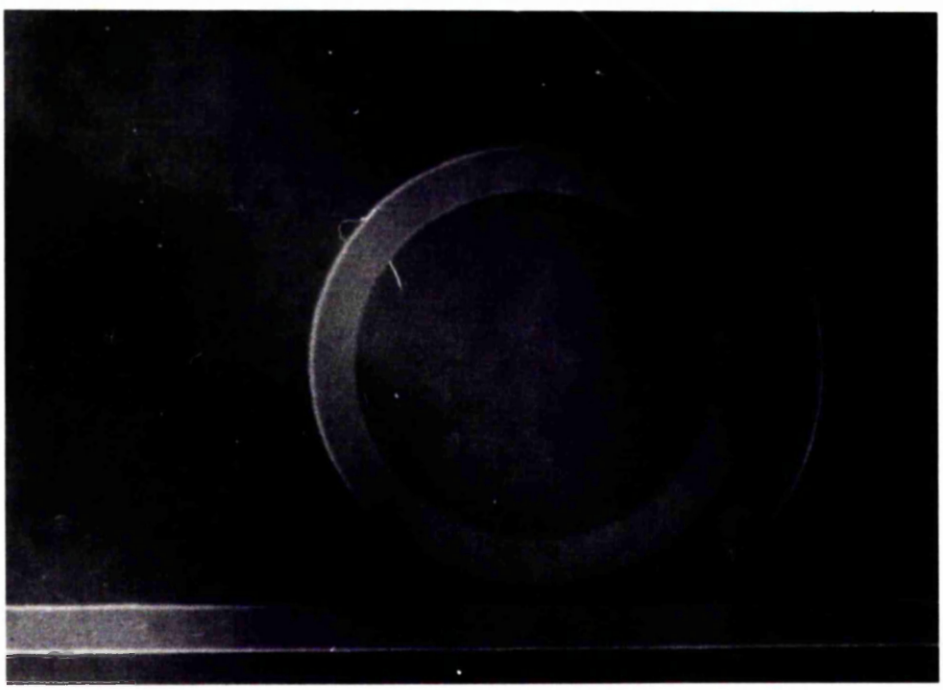
This ring pattern was chosen purposely to show up the following phenomena, which are obviously closely akin to the association effects noted in (3) above.

(a) When two waveguides converge to within about 8 μm of each other, the discolouration along their inner edges rapidly fades and vanishes.

(b) If a piece of the aluminium mask is electrically isolated from the bulk (e.g. the ring centres), then the amount of discolouration at an edge defined by this will depend upon its size. For example the smallest ring is devoid of reduced silver on its inside boundary; the



(a)



(b)

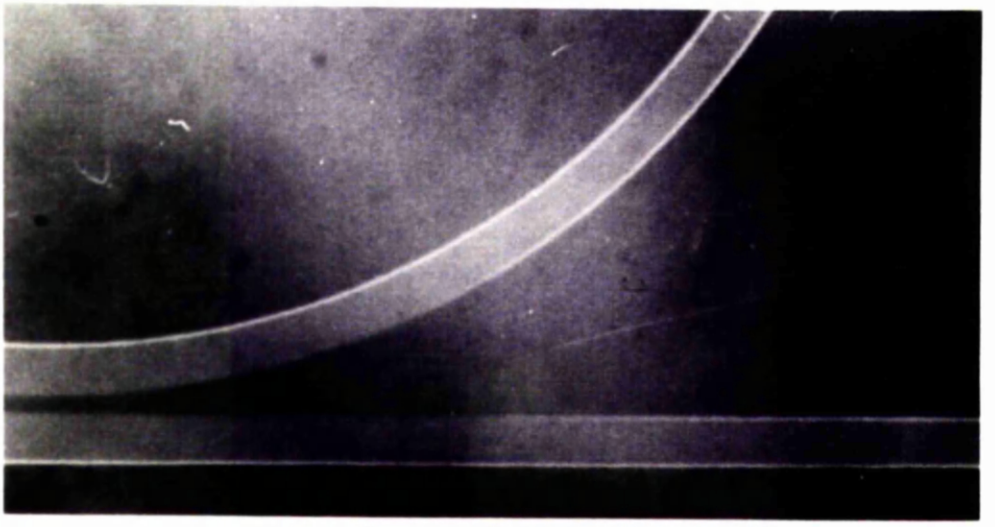
Fig. 8.9 Photo-micrographs of the ion-exchanged surface:

(a) half metallic, half anodised diffusion mask.

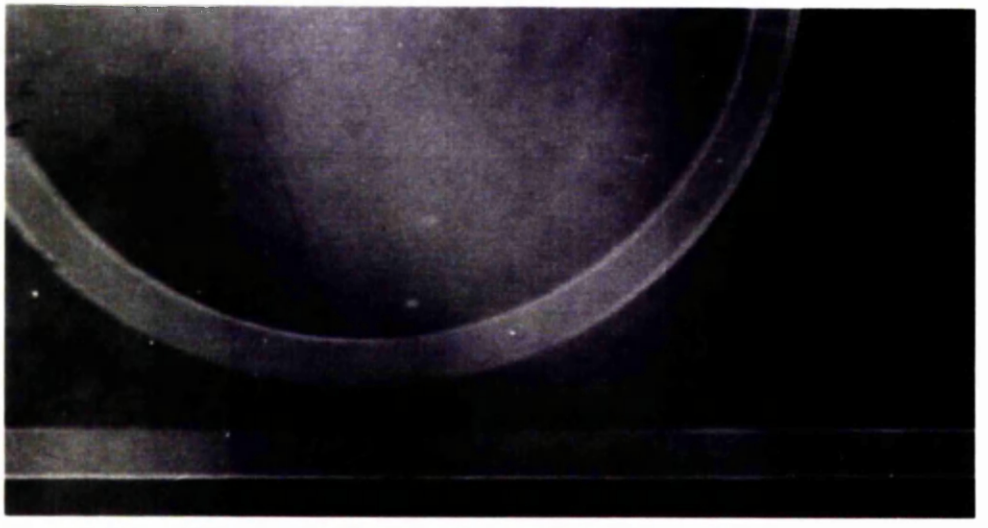
(b)-(f) (overleaf) metallic diffusion mask.

Ring diameters:

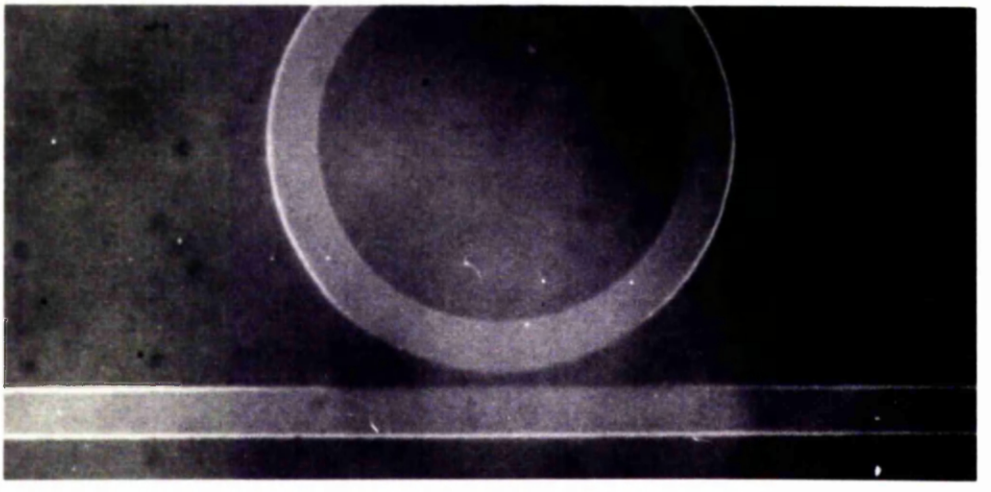
(c) 200 μm, (d) 100 μm, (b,e) 50 μm, (f) 25 μm



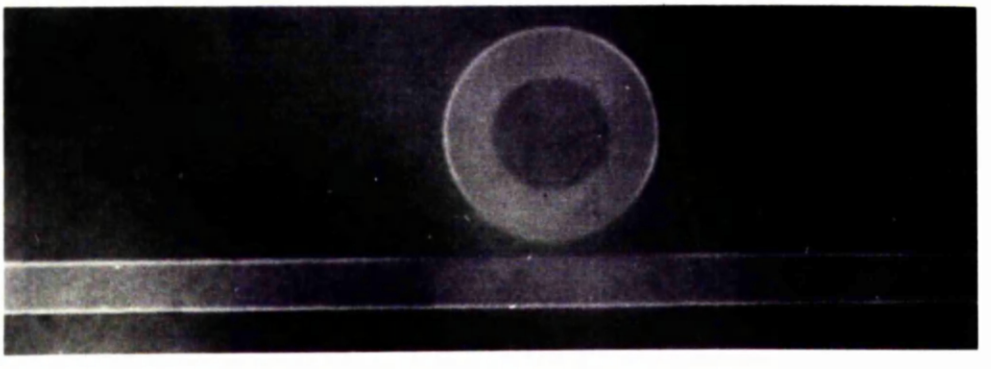
(c)



(d)



(e)



(f)

next smallest has a trace; the two largest have a considerable, but varying, amount.

Waveguides made using an anodised diffusion mask are unconditionally free from reduced silver. Also shown in Fig. 8.9 is the surface of a half anodised, half metallic diffusion mask; the discolouration stops abruptly at the transition point. As may be expected, partially anodised diffusion masks produce a lower density of reduced silver than non-anodised.

Because of this, metal-mask waveguides are found to have a much higher attenuation than anodised-mask guides, for which reason metal masks are no longer used.

8.4 Discussion

It is clear, from the foregoing evidence, that the cause of these strong deviations from the ideal behaviour in the two extreme cases, is electrical in nature. The origin and mechanism of this will now be considered in detail. Since metallic masks are no longer used and good insulating masks can now be made, no attempt has been made to incorporate the following considerations into the mathematical model. The discussion will be mainly qualitative.

The interdiffusion theory of Chapter 4 is believed to be accurate within the bounds set. A case is made, therein, for the belief that the mask conductance has little intrinsic effect on the final diffusion profile and cannot directly affect the quantity of silver taken in through an aperture of fixed size. The grounds for this are thermodynamic. The

diffusion rate is fixed by the free-energy of the system and can be enhanced or retarded only by an input of energy from some external source.

1) The Metallic Mask

In this case, the only available, external energy source is electrochemical in nature. It is very probable that an electrolytic cell has been inadvertently set up, with aluminium and glass electrodes in an electrolyte of molten silver nitrate. The glass electrode potential (see section 4.8) results from the ion-exchange itself. We may consider the ion-exchange to be initiated by an out-flow of sodium from the glass (the sites must be vacated before silver can replace it); the resultant charge separation and glass-to-melt potential difference assist the in-flow of silver. Thus the glass is negative with respect to the melt.

The aluminium electrode potential arises by a mechanism which is analogous to the above. Dissolution of the metal in the electrolyte causes a charge separation, since Al^{3+} ions enter the electrolyte, leaving three electrons behind. This encourages silver to deposit on the metal surface. At equilibrium:

$$Al \rightarrow Al^{3+} + 3e^{-}$$

$$3\,Ag \leftarrow 3\,Ag^{+} + 3e^{-}$$

Thus the metal is also at a negative potential with respect to the electrolyte.

Three resistivities are of interest.

(i) Glass: using the values of α and D_a found in Chapter 7.

$$\rho_g \simeq 10^{11}\ \Omega\ \mu m \qquad \text{at } 215°C.$$

From a suitable data book;[63]

(ii) Melt: $\rho_m \simeq 1.5 \times 10^4$ Ω μm at 215°C

(iii) Aluminium: $\rho_a \simeq 40$ Ω μm

The aluminium is in the form of a thin film of about 0.1 μm thickness; therefore the relevant parameter is a sheet resistance:

$$R_s \simeq 400 \ \Omega$$

The glass and aluminium electrode potentials may be considered as voltage sources, (v_g and v_a), in an equivalent circuit; their positive terminals connected via a melt resistance (R_m) and their negative terminals, via a glass resistance (R_g). If the two voltages differ current will flow and, since $R_g >> R_m$, most of the potential difference, $v_a - v_g$, occurs between the metal and the ion-exchanging glass surface. Unless this current flows, no work can be done to either enhance or retard the diffusion.

We consider the two possibilities for the sign of $v_a - v_g$:

(a) $v_a < v_g$. If the aluminium is the positive terminal then current flow may occur either by electron flow through the glass to the metal, or injection, by the metal, of Al^{3+} ions into the glass. Both of these mechanisms appear to be highly improbable since glass has no electrons available for conduction and aluminium, when present in glass, is never mobile but is bound into the lattice. If this polarity occurred, the imposed field would be neutralised by the establishment of a region, near the junction, depleted of positive ions and no further current would flow.

(b) $v_a > v_g$. Here the aluminium is the negative terminal, and current flows by the reduction of cations (which drift towards the mask) by electrons which the metal supplies. Thus, the reduction of free, metallic silver or sodium is an essential consequence of current flow if the mask is metallic.

The evidence [see 8.3 (4)] confirms that this is indeed the case, and it is concluded that $v_a > v_g$. It is unlikely that any sodium can be reduced by a potential, generated by the ionisation of aluminium, which is less reactive than sodium. Thus, especially since silver is always available for the purpose, it may be assumed that the current is carried solely by the reduction of silver at the mask edge.

In effect, silver is withdrawn by the mask, since metallic silver is very much more concentrated than the ionic silver in the glass, and has little influence upon the refractive index.

In the light of the above deductions, we consider phenomena associated with the silver reduction (see Fig. 8.9). It is noted that only the edge of the mask, where silver is present beneath, is supplying electrons to the glass (since sodium is not reduced), but that these electrons will be drawn from a comparatively large area of aluminium. It is observed that when an isolated area of aluminium is small (less than about 30 μm diameter) it is incapable of supplying sufficient electrons to its edge. Two explanations are possible.

(i) The aluminium electrode potential may be weakened by proximity to ion-exchanging glass, either due to capacitive effects or to electrolyte modifications. When the area is large, the regions far from the edge hold down the potential and supply electrons.

(ii) The voltage source represented by a unit area of aluminium may be non-ideal, having internal resistance, due to the melt and the aluminium resistivities, or to the limited rate of the oxidation reaction which generates the voltage. For an isolated, circular region of radius r, the ratio of perimeter to area is 1/2 r. Thus as the radius is reduced, the current drawn from each unit voltage source increases, the effective electrode potential is weakened and the current per unit length of perimeter is reduced.

Of these two hypotheses, (ii) is the more satisfactory. A narrow strip of metal between two parallel guides is similarly unable to supply many electrons; they must be donated by the open mask area beyond the point where the guides diverge. However, the current path is resistive (a 4 μm wide strip has a resistance of 100 Ω per micron of length) and is itself a consumer of electrons, so the current decays rapidly with distance. Thus, the narrow strip behaves as though isolated.

The Width Dependent Retardation

When current flows in an electrolytic cell consisting of two dissimilar metals, dissolution of the more reactive metal is accelerated and that of the other metal is retarded. This is the principle of cathodic protection against corrosion, using a sacrificial anode.

In this case the cathode is an ion-exchanger, but the principle still holds; the thermodynamic process which sets up the cathode potential (out-flow of sodium) is retarded and that which responds to the potential (in-flow of silver) is accelerated: both rates are changed. Thus, if in one second with no current flow the glass loses n sodium ions and gains n silver ions per unit length of guide, then in the same time, with current flow, the sodium loss is n - p ions and the silver gain is n + q ions.

p and q are not independent, but vary together. The net material gain through the glass surface is p + q ions and a mean current of (p + q)e Amps per unit length of guide has flowed through the glass surface and into the mask edges.

In order to make room for this extra material in the glass a total of p + q silver ions (exclusively silver) are reduced and deposited beneath the mask edge. This silver has effectively been withdrawn from the system, so the real gain of silver (in one second, per unit length of guide) is n + q - (p + q) or n - p ions. The effective ion-exchange rate has been reduced by the current flow.

The overall rate reduction, p ions/second/unit length, depends only upon the total current and is independent of the aperture width. However, the effect upon the average depth of the waveguide is given by the mean rate-reduction per unit area, which is p/a, where a is the aperture width (the mean current density at the glass surface also varies as $1/a$).

To check that the observed width dependent retardation follows the above relationship, we consider the data given on p.117. In order to model the reduced ion-exchange rate, a reduced surface silver concentration (normalised), $u_0(a)$, was used in DIF. If it is assumed that u_0 gives an approximate measure of the mean ion-exchange rate (ions/second/unit area), then the rate-reduction, for each width, is $1 - u_0(a)$. Thus, if the above hypothesis is correct:

$$1 - u_0(a) \propto p/a$$

or $$[1 - u_0(a)]a = K; \quad \text{where } K \text{ is a constant.}$$

From the data on p. 117:

$1 - u_0$	a(μm)	K
0.11	2.95	0.325
0.075	4.35	0.326
0.05	5.98	0.299

The approximate constancy of K gives considerable support to the above hypothesis.

About the profile shape little can be said, except to note that the electric field, in moving ions from the centre toward the edges, will lessen the effective width reduction which the retardation (greatest near the edges) would otherwise cause.

It is concluded that the hypotheses of this section can explain all the phenomena associated with an aluminium diffusion mask. The partial anodisation results are probably caused both by weakening of the electrode potential and reduction of the mask conductance. The "association" effects [8.3 (3)] are due to geometrical limitation of the electron source and are akin to the silver deposition phenomena noted in section 8.3 (4).

2) The Anodised Diffusion Mask

With an ideal, insulating diffusion mask current flow, with consequent modification of the diffusion rate, is impossible. Such a mask isolates the glass, which it covers, from the melt; the "circuit" is "open" and any potential between the mask and the melt is irrelevant.

However, since with the early anodised masks enhancement of the diffusion clearly occurs, it may be deduced that the mask is not ideal but permits ionic current flow. (The possibility of electron flow may be discounted immediately). This being established, we proceed to use the explanation of the metal-mask behaviour in a slightly modified form.

It is postulated that the oxide film has selective ionic diffusion properties, allowing sodium to pass through much more readily than silver. Thus, out-flow of sodium sets up both the potentials v_a and v_g, however v_a, unlike v_g is not reduced by an influx of silver. Because of the ionic filtering properties of the oxide, v_a is the electrical equivalent of osmotic pressure.

If this is so, then as before, $v_a > v_g$ and the current flow situation is exactly as described above (p. 127). The only difference is that the mask extracts charge by means of the exclusive withdrawal of sodium ions rather than by the exclusive reduction of silver ions. In one second, n + q silver ions are gained and n - p sodium ions lost through the glass surface. The excess (p + q) ions are extracted as sodium via the mask, thus the net sodium loss is n - p + (p + q) or n + q ions. The net silver gain is also n + q ions instead of n, as it would have been with no current flow. Thus the effective ion-exchange rate has been increased, especially near the edges of wide waveguides. The edge-deepening will be emphasised by the tendency of the field to draw silver towards the edges, resulting in considerable widening.

That the anodised film behaves in this way cannot be stated with certainty, but is quite probable. The non-porous barrier layer at the base of the oxide film is only about 150 Å thick (see section 6.2) and either aluminium of oxygen atoms (or both) diffuse through it during the anodisation.

As was stated earlier, the experimental factor which causes the width-dependent enhancement effect has not been isolated. In the light of the above theoretical discussion (if the hypothesis is true), the

various anodised mask characteristics should reflect varying conditions of the anodised barrier layer: thick or thin, hydrated or otherwise.

Assuming that the aluminium-to-glass adhesion is good, the final waveguide characteristic should depend largely on how the anodisation is carried out. In particular, the post-anodisation rinse may be quite critical; an overlong soak in slightly acidic water at room temperature could seriously degrade the barrier layer by hydration and dissolution. The evidence is not complete, but it is known that some, at least, of the improved anodised masks were placed in a bath of acetone after anodisation (after a brief aqueous rinse); later masks, which were soaked in water after anodisation, proved somewhat permeable to silver ions.

An alternative anodisation system, using tartaric acid (partially neutralised with ammonium hydroxide) and a very high voltage (greater than 100 V) produces a dense, non-porous oxide film, of thickness ~ 12 Å per volt.[58] This is worth investigation as a means of producing an ideal insulating diffusion mask.

8.5 Post-Baked Stripe Waveguides

Three sets of stripe waveguides, (with improved, anodised masks), were diffused, for 7, 10 and 20 minutes (the initial diffusion time, t). The dispersion characteristics were measured, and were found to follow the computed theory fairly closely (see Fig. 8.3). They were then post-baked (section 6.4) for times 2t and the modes measured before post-baking for further times of 6t, a total post-bake time of 8t. The three resulting sets of mode spectra are presented in Figs. 8.10, 8.11 and 8.12 for the 7, 10 and 20 minute diffusions, respectively. There are no theoretical curves in these figures.

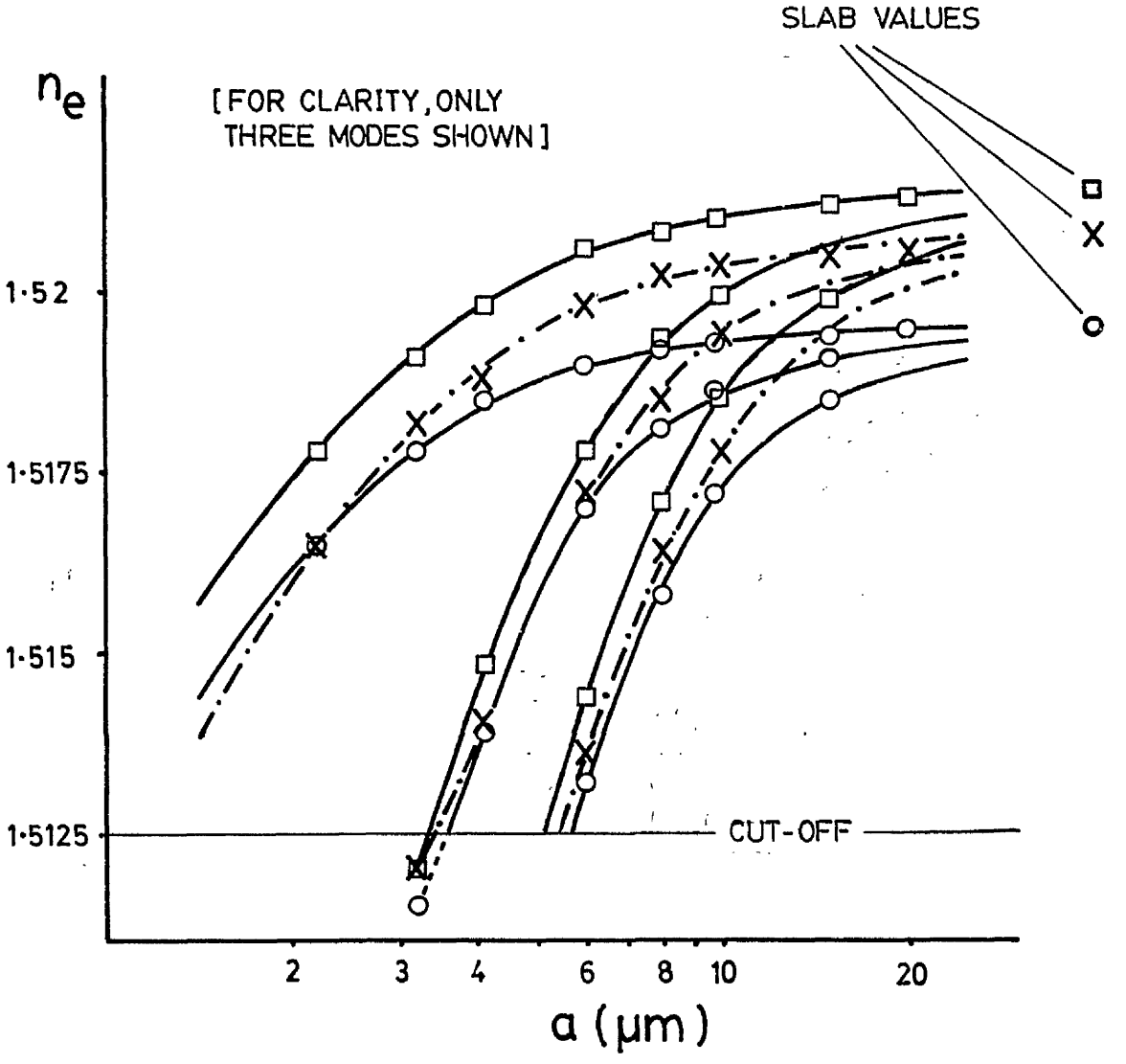


Fig. 8·10 Experimental graph of n_e vs. aperture width [a] for post-baked, stripe, ion-exchanged waveguides.

—o— initial diffusion t = 7 minutes

—□— 2t post-bake

—·—x—·— 8t post-bake

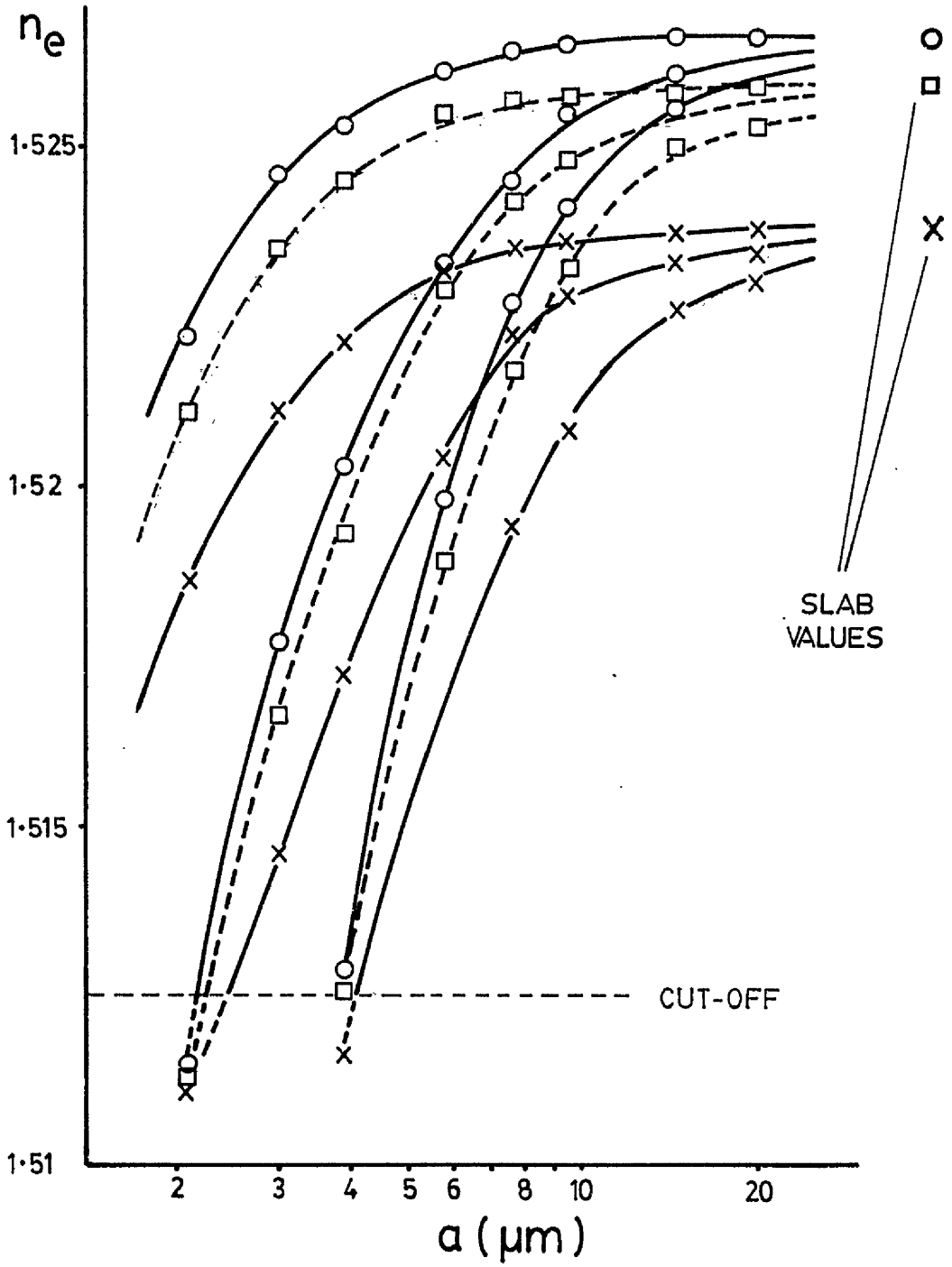


Fig. 8·11 — As Fig. 8·10

—o— initial diffusion t = 10 minutes

---□--- 2t post-bake

—x— 8t post-bake

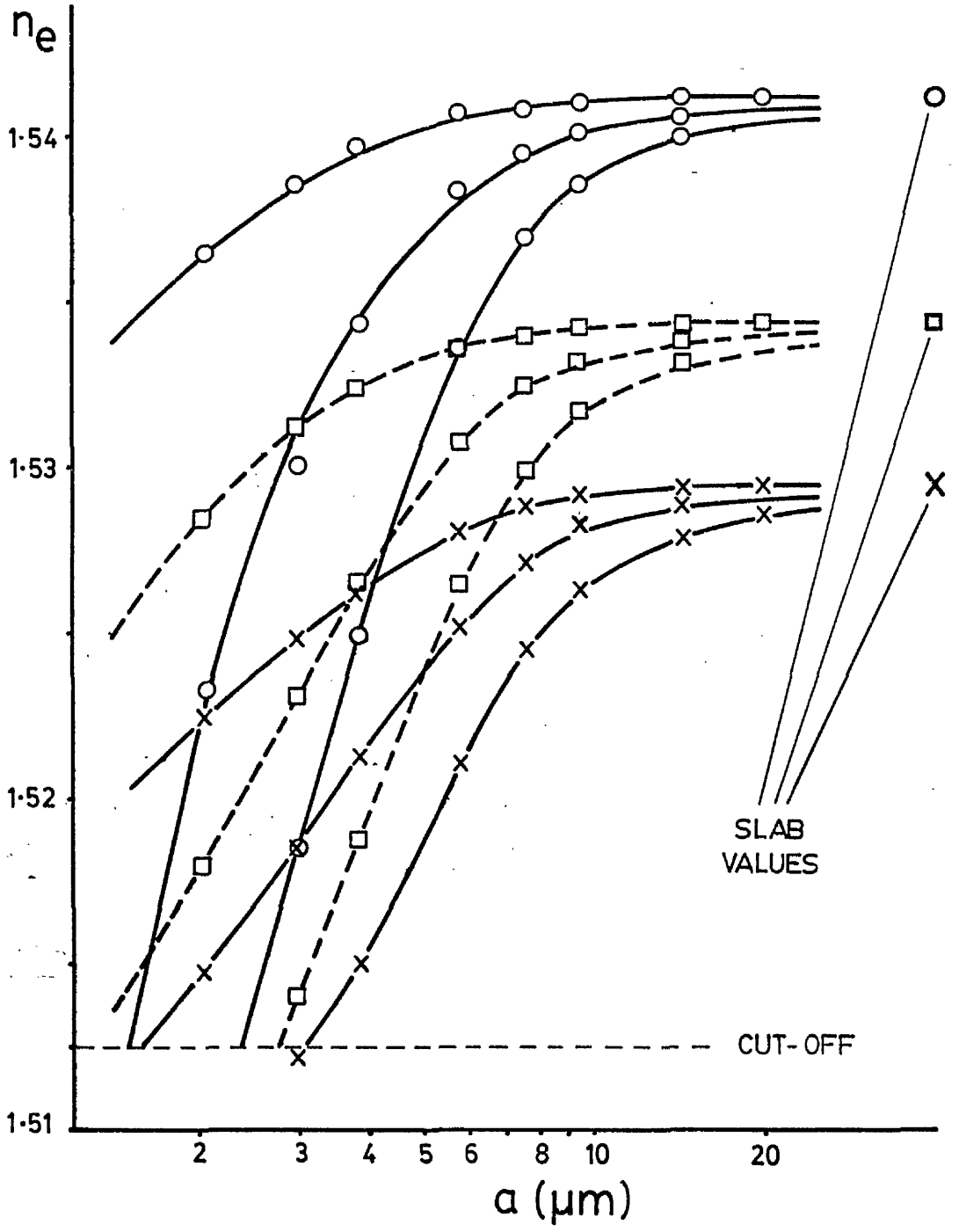


Fig. 8·12 — As Fig. 8·10

—○— initial diffusion t = 20 minutes

---□--- 2t post-bake

—×— 8t post-bake

The main features of these results may be summarised as follows:

(i) The stripe effective indices usually move in the same direction as the slab value during post-baking (see Fig. 7.9), but the average width-dispersion (i.e. the slope of the curve) tends to increase. The fundamental-mode n_e for the narrowest stripe moves down more rapidly, or up less rapidly than the slab value.

(ii) Modes near cut-off move (up or down) less rapidly than well-guided modes. Thus if the slab n_e decreases the mode-separations decrease, and vice-versa.

As was decided earlier, we require a single-mode waveguide whose effective index is not too low. (How low a value is tolerable will be considered in Chapter 9.) Thus we are primarily interested in waveguides whose widths lie between 2 and 3 microns; 2 microns is the approximate lower limit for guides made using standard photolithography.

Clearly, 20 minutes diffusion is too long. Because the mode-separation decreases, no practical amount of post-baking seems likely to make a 2 micron guide single-mode . The 7 minute diffusion seems to be unconditionally single-mode for widths between 2 and 3 microns. Probably, no amount of post-baking will make the second mode appear. (Note that at 3.4 μm, this mode rises through cut-off, then falls below it again as the post-baking proceeds.) However, n_e is rather low.

A 10 minute diffusion allows a maximum width of about 2.2 μm for single-mode guidance, suggesting that the initial n_e values of 1.523 for the stripe, and 1.5266 for the slab are close to the maximum permissible values. The evidence suggests that all modes move in the same direction during post-baking, therefore we require the second mode (n = 1), after the initial diffusion, to be at, or just below, cut-off. The post-bake time (if any) must be sufficient to ensure that all n_e values either decrease or remain stationary, as indicated by the change in the slab n_e.

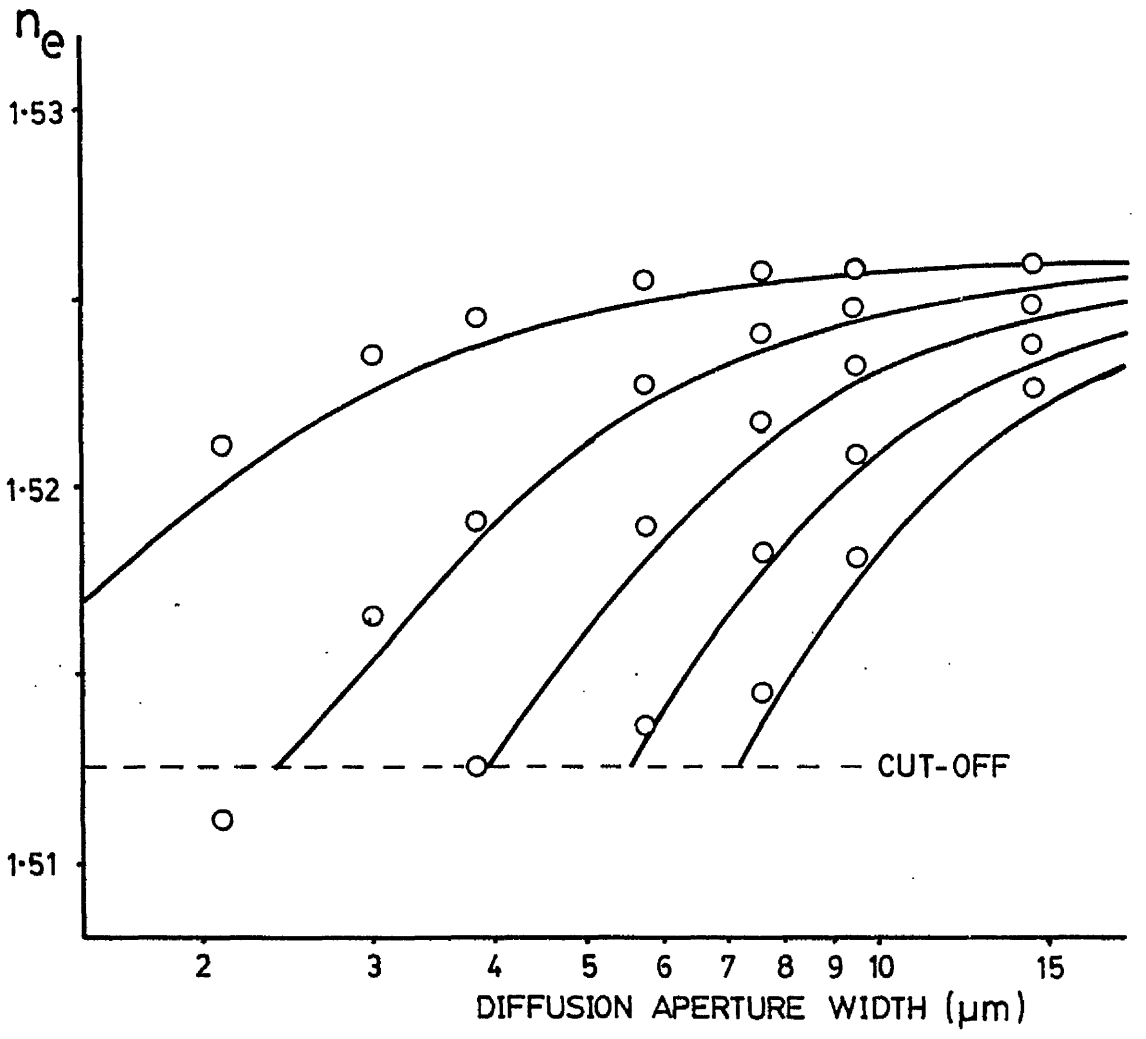


Fig. 8.13 Comparison of experimental [o] and computed [——] modes for post-baked, stripe waveguides.

initial diffusion time t = 10 min.
post-bake time = 2t

In Fig. 8.13 the computed and experimental dispersion curves are compared for t = 10 minutes and a post-bake time of 2t. In order to obtain a true comparison of the lateral-mode behaviour, the two computed slab values (initial and post-bake) are artificially matched to the measured values. This is found to require a computational post-baking time of $1.81t_n$ rather than $2t_n$ (where t_n is the computational diffusion time).

As may be seen, there is a tendency for the narrow guides to remain deeper, and the wide guides wider than the theory, as with the initial diffusion (similar to the 11 minute diffusion of Fig. 8.3). This is probably due to the residual width dependent enhancement of the improved anodised mask; but there may also be a problem concerning the concentration dependence of the diffusion rate, as noted with slab guides (see Chapter 7). The discrepancy probably becomes worse with long post-bake times.

8.6 Summary of Conclusions

In this chapter, experimental results for stripe waveguides made using aluminium and anodised (aluminium oxide) diffusion masks have been presented. The nature of the discrepancy between these and the computed theory suggests that electro-chemical potentials drive current through the glass, either accelerating (anodised mask) or retarding (metal mask) the ion-exchange. Reduced silver metal beneath the edges of a metal mask makes this unsuitable for use, so the mask must be anodised. In this case, ionic conductivity of the anodised mask may cause problems since the barrier-layer of the oxide is very thin. This can be overcome (the improved anodised mask) in which case the measured and computed waveguide modes are in good agreement. More work is required to improve the anodised mask further, and it is suggested that the tartaric acid/ammonium tartarate anodisation process (which produces a non-porous oxide film) be investigated.

In order to obtain single-mode, stripe waveguides of width 2 μm, the diffusion time must be limited to 11 minutes or less. Slightly longer times (up to 13 minutes perhaps) are permissible if post-baking is used to lower the second mode n_e below cut-off.

CHAPTER 9

EXPERIMENTAL RESULTS FOR THE RING-RESONATOR ELEMENTS

In this chapter, experimental results for waveguide bends, directional couplers and the straight-guide attenuation will be presented. These provide the necessary, basic information for the design of a ring-resonator. After discussing the results, they are brought together into just such a design.

All the devices considered in this chapter use the same input coupling technique (mentioned, but not described in detail in Chapter 6) so this will be considered first.

Waveguide Excitation

In order to maximise the guided light intensity, tapered waveguide sections are used to concentrate the energy into the narrow (2-3 μm) stripes. Mode measurements are not required, so a short focal-length lens, about 3 cm from the coupling prism, is used. This distance has a fine adjustment so that the beam-neck (less than 20 μm diameter) may be accurately positioned, with the aid of a microscope, at the taper entrance. The light is launched into a slab waveguide area so that this adjustment may be readily carried out; the input prism is placed well back from the taper entrance to permit ready access, to the stripe, of the output prism or microscope objective.

9.1 Straight Guide Attenuation

The Mask Pattern

The straight-guide attenuation is estimated by measuring the intensity of the light from an output prism at various positions on the guide. Obviously, this requires that a reasonable length of guide be available, preferably at least 2 cm. However, the aspect ratio of such a pattern (2 cm : 2 μm = 10^4) is so large that considerable difficulties are encountered in the cutting and photo-reduction processes which are required to produce the shadow-mask (see section 6.1).

A suitable, commercially made shadow-mask was found to be available. The pattern consists of several guide-sets, each with four waveguides which converge to form coupled sections at various intervals. On two of the guide-sets, the guide separation at these coupled sections (4 μm) is sufficient that no interaction will occur; the nominal guide-widths are between 2.5 and 3 μm.

Both outer guides of each set are perfectly straight, allowing these to be used for attenuation measurements, but the inner guides have stepped sections (since the pattern is computer-generated) to converge or diverge the guides. These are found to cause considerable attenuation. Other guide-sets have wider waveguides - about 6 μm. The available length is about 2.5 cm (not including the taper) and the edge quality is good.

The Measurement Technique

Having maximised the guided light intensity, as judged by the surface scatter, the output prism is clamped on to the far end of the waveguide, tightening the screw until the guided light stops abruptly just within the prism. There should be no bright, scattering point at the prism corner.

The output m-line is extensive and must be condensed by means of lenses. A cylindrical lens, close to the prism reduces the diffraction angle before the m-line becomes too long, and two 4 inch focal-length lenses in contact then focus the light on to a large-area photo detector. The laser beam is modulated at 1 kHz by means of a "chopper" and a 1 kHz, tuned amplifier (a Pye V.S.W.R. meter) with up to 40 dB of attenuation (in 1 dB steps) is used to measure the intensity. Various screens are necessary to deny stray, scattered light access to the lenses.

After each measurement the distance from the edge of the output prism to some reference point is measured to the nearest 0.25 mm.

Usually, at least five readings are taken per stripe, moving the output prism towards the input prism. The prism is sometimes found to damage the waveguide, so the measurement cannot be repeated. Similarly, the attenuations of the same guide, before and after post-baking, cannot be compared directly. The post-baked sample must be unused.

Interpretation

There are two sources of error in these measurements:

(i) variation in the output coupling efficiency. This is as likely to reduce the apparent attenuation as to increase it.

(ii) the presence of large flaws in the waveguide, which can only increase the apparent attenuation. Since it is quite possible for a small ring (say 500 μm in radius) to be free from such flaws, these are not considered to be representative.

For each waveguide, the guided intensity (in dB) is plotted against distance (in cm). The points are connected by straight lines, with slope α_L dB/cm, and discrete vertical intervals, representing defects. Sometimes these jumps correspond to visible scattering centres, sometimes not;

it is found that only surface defects produce a bright scatter spot. On several samples, the same guide was almost interrupted due to dirt on the original shadow mask; a discrete interval was observed in the measured intensity at each side of this defect, yet no bright spot was visible. Probably light is scattered through the sides of the guide but remains invisible because of total reflection from the glass surface.

Thus the value of α_L is biased towards the shallower gradients measured, see, for example Fig. 9.1.

Experimental Results

A large batch of long, stripe waveguides (15 slides) was made. All diffusion masks, except one, were anodised and, by utilising different guide-sets, using MF312 developer and over-developing some samples, guide widths ranging from 5.8 μm to 1.7 μm were obtained. All were diffused for 9 minutes; some were not post-baked, some were post-baked for 9 minutes (t) and some for 27 minutes (3t).

In Fig. 9.2 the linear attenuation constant, α_L, (dB/cm) is plotted against the reciprocal of the aperture width [following C.S. Wilson;[61]] an approximately linear relation was expected.

It is seen that α_L increases as the width diminishes and that post-baking reduces α_L. Significantly, the first 9 minutes of post-baking are of greater benefit than the next 18 minutes, a fact which is probably due to the rapid relaxation of gradients, with consequent edge smoothing, which occurs during the initial stages of post-baking.

Thus, according to this data, one unit of post-baking (i.e. t minutes) with a width of 2 μm allows a linear attenuation of 2.25 dB/cm to be achieved. The slab waveguide attenuation (with no post-bake) depends

upon the surface quality, but can be as low as 0.4 dB/cm. It is not strongly dependent upon the diffusion time.

In an attempt to remove the gross surface imperfections, several slides were ground and polished to an optical finish (on one side only). This resulted in a large increase in the surface scatter with a consequent increase in α_L of several dB/cm. However, fewer gross defects were noted; thus the overall attenuation was more predictable than with unpolished slides. The samples of Fig. 9.2 were from a different batch of Fisher slides and were found to be remarkably defect-free by comparison with the batch from which the polished slides were taken.

The metallic-mask waveguide, which was about 3 μm wide had an attenuation, α_L, of 7 dB/cm.

Unfortunately, the validity of the data of Fig. 9.2 is not unqualified. The guides were intended to be single-mode at 2.3 μm width, yet all (including those less than 2 μm wide) are found to be double moded. On measuring the modes, the fundamental n_e was found to be little less than the slab value and the mode separations much smaller than expected from the aperture widths; smaller even than those of the "early anodised" characteristic.

It seems likely that these anodised masks were partially permeable to silver ions, resulting in a reduced lateral index difference. Permeable masks, which can be modelled using DIF, have the effect of making the guides seem wider. The cause of this probably lies in the post-anodisation rinsing; after a brief rinse, the slides were placed in a bath of distilled water. It was not suspected, at that time, that this might damage the barrier layer.

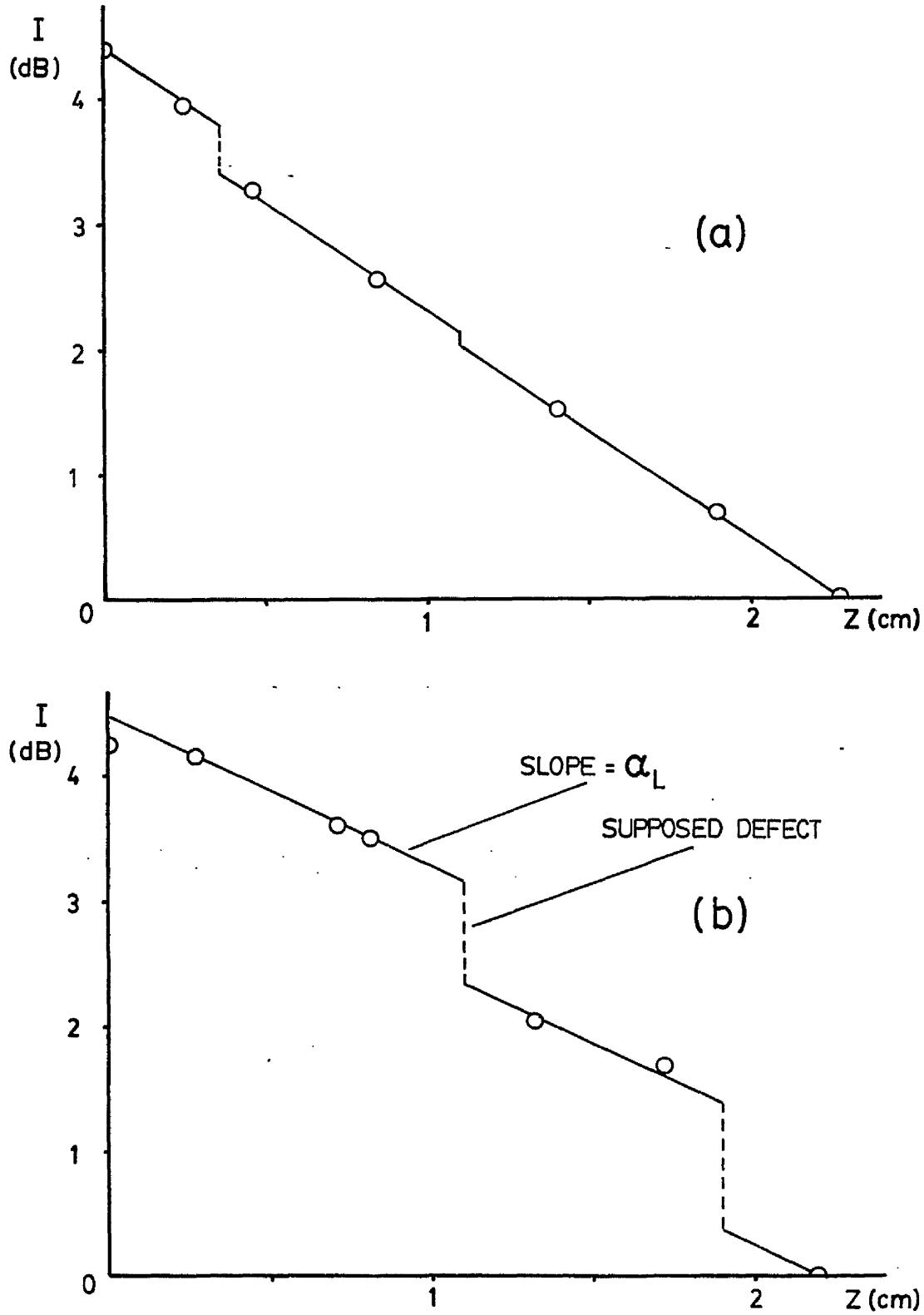


Fig. 9.1 Experimental graphs of guided light intensity[I] vs. propagtion distance[z]

These interpretations give α_L values of:
(a) 1·75 dB/cm ; (b) 1·2 dB/cm

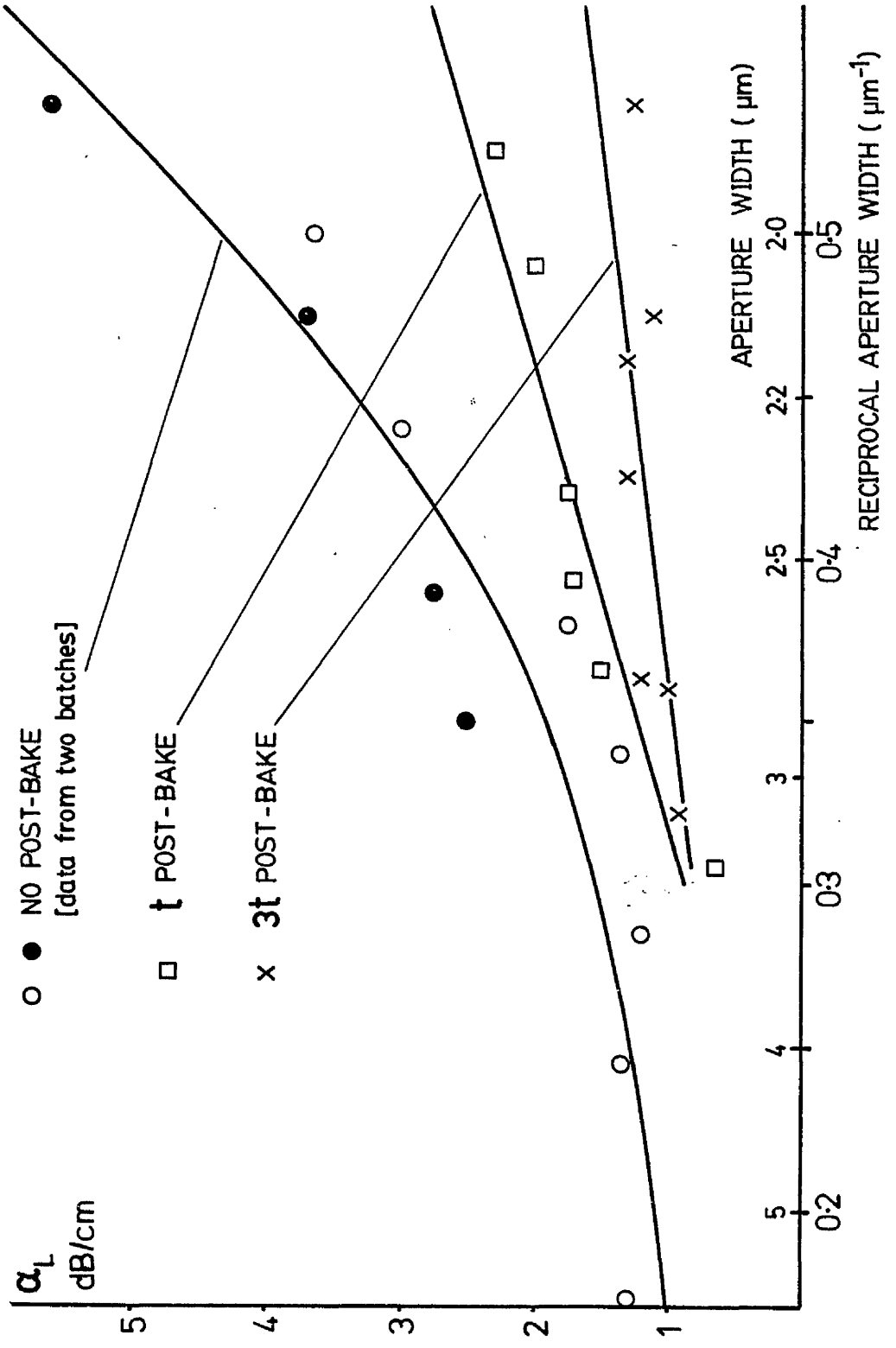


Fig. 9·2 Measured linear attenuation factor [α_L] plotted vs. reciprocal width.

diffusion time, t = 9 minutes

One effect of this may be to decrease the attenuation. The points on Fig. 9.2 marked as filled circles (no post-bake) are from an earlier batch of samples. These were truly mono-mode and thus may represent better quality masking. The attenuation is slightly higher than for the other samples, thus it may be safer to assume an attenuation of about 3 dB/cm at 2 μm width after one unit of post-baking.

Clearly, this figure could be improved by the use of better substrate material. Optically clear, uncoloured glass (Fisher slides are green) with a flawless molten surface could, in theory, be produced; and probably would be if these guides were made commercially.

9.2 Directional Couplers

The pattern used for making directional couplers is illustrated in Fig. 9.3. The waveguide width (a) and separation (c) are nominally 2 μm, however these may be varied from a = 1.92, c = 1.85 to a = 2.6, c = 1.3 microns by changing the exposure and developing times. Three interaction lengths (s), of 1 mm, 2 mm and 4 mm, are available.

The measurement method is similar to that described in section 9.1. Light is coupled into the straight guide (guide 1) via a taper and a certain fraction transfers to guide 2 before it diverges from guide 1. The output prism is clamped on with its rectangular edge close to the divergence point, resulting in two spatially separate m-lines whose intensities may be measured by moving the condenser lens and detector up or down. If the intensities in guides 1 and 2 are I_1 and I_2, then:

$$\frac{I_2}{I_1} = \frac{\sin^2 (Ks)}{\cos^2 (Ks)} = \tan^2 (Ks) \qquad (9.2.1)$$

where K is the coupling coefficient and s is the interaction length.

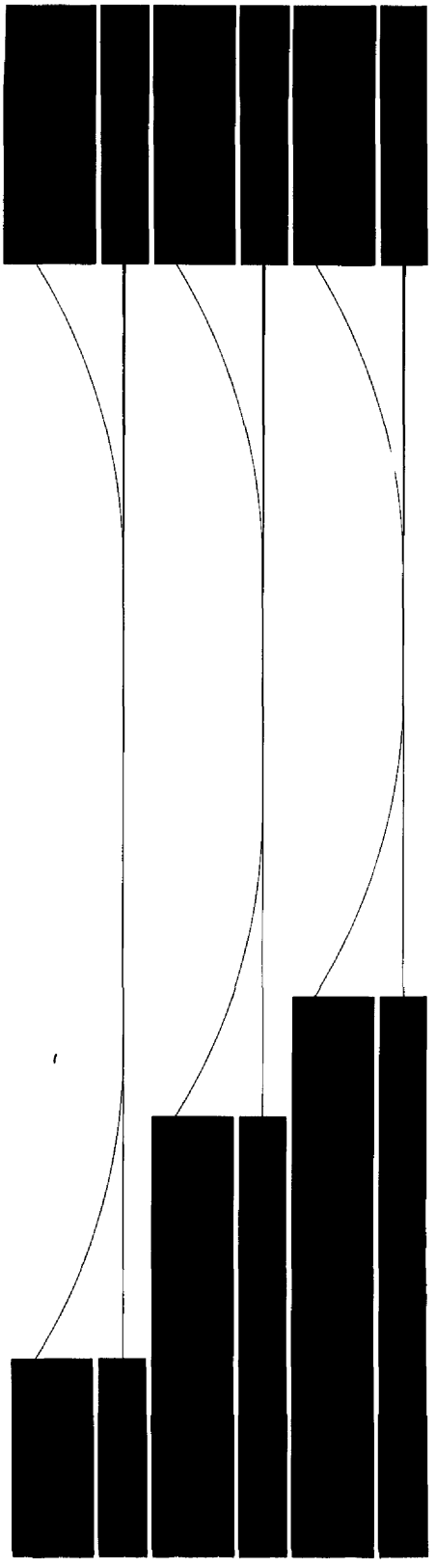


Fig. 9.3 The directional coupler shadow-mask

The transfer length, $L = \pi/2K$.

Since it is not possible to tell how many times the energy has passed between the guides, the general solution for Ks is:

$$Ks = m\pi \pm \tan^{-1}\left[\frac{I_2}{I_1}\right] \quad : \quad m = 0, 1, 2 \ldots . \qquad (9.2.2)$$

where the negative solution for m = 0 is discarded.

With the interaction lengths available here, only m values of 0, 1 or 2 are of practical importance. The criteria used to judge the correct value of L will be considered later.

Results

Several coupler samples have been made, each slide containing two coupler-sets (Fig. 9.3), i.e. six couplers, two of each length.

The first successful device, which was made at a time when the "improved anodised" mask seemed to be stable and reproducible, was diffused for 9 minutes, resulting in a slab n_e of 1.5253. Accordingly, this value was made the basis of a comparison between computed and measured transfer lengths.

Since most of the couplers on this slide had dimensions in the vicinity of a = 2.3 μm, c = 1.5 μm, a preliminary computation, using these values, was carried out to check the effect of the parameter α. With $\alpha = 0.7$ and $n_1 = 1.599$ (a = 2.32, c = 1.493) a computed transfer length (L) of 4.46 mm was obtained. With $\alpha = 0.9$ and nothing else changed except the computational diffusion time, t_n, which was adjusted to return the slab n_e to 1.5253, L increased to about 8 mm.

Since the measured transfer lengths for these couplers lie between 4.13 and 5.44 mm, α was kept at 0.7.

It is not generally possible to include the exact, measured, coupler dimensions in the computer model. The diffusion aperture must be an integer multiple of δx and the aperture separation, of $2\,\delta x$ (because of the lateral symmetry). δx and δy - the refractive index sampling intervals - are both about 0.04 μm, thus the accuracy of the model is limited, especially since L is so sensitive to the separation (c).

To overcome this difficulty, a set of 16 coupler computations were performed, using four widths and four separations. The transfer length of any coupler may then be found almost exactly, by interpolation. Fig. 9.4 presents this computed data as a family of four curves, L vs. a with c as the fixed parameter for each curve. In order to perform the interpolation, this must be converted into another family of curves, L vs. c, one curve for each value of a required; this working graph is not shown.

Finally, the comparison of the computed and measured transfer lengths is contained in Table 9.1. (Only five couplers are considered because coupler A_2 was visibly imperfect.) The most probable measured value of L is selected as being the value which is most consistent with those obtained using different interaction lengths (the short couplers are useful here). If this still leaves any doubt, the value which is closest to the theory is chosen. The choice, in this case, is not difficult as is illustrated by the column of alternative measured values.

The measured and computed values of L are, for the most part, in good agreement. In three cases out of five the error, expressed as a percentage of the computed value, is less than 10% and in only one case is it above 25%. It is considered that, despite the small sample, the agreement is significant and not due to chance. The causes of the, occasionally large, error will be considered later.

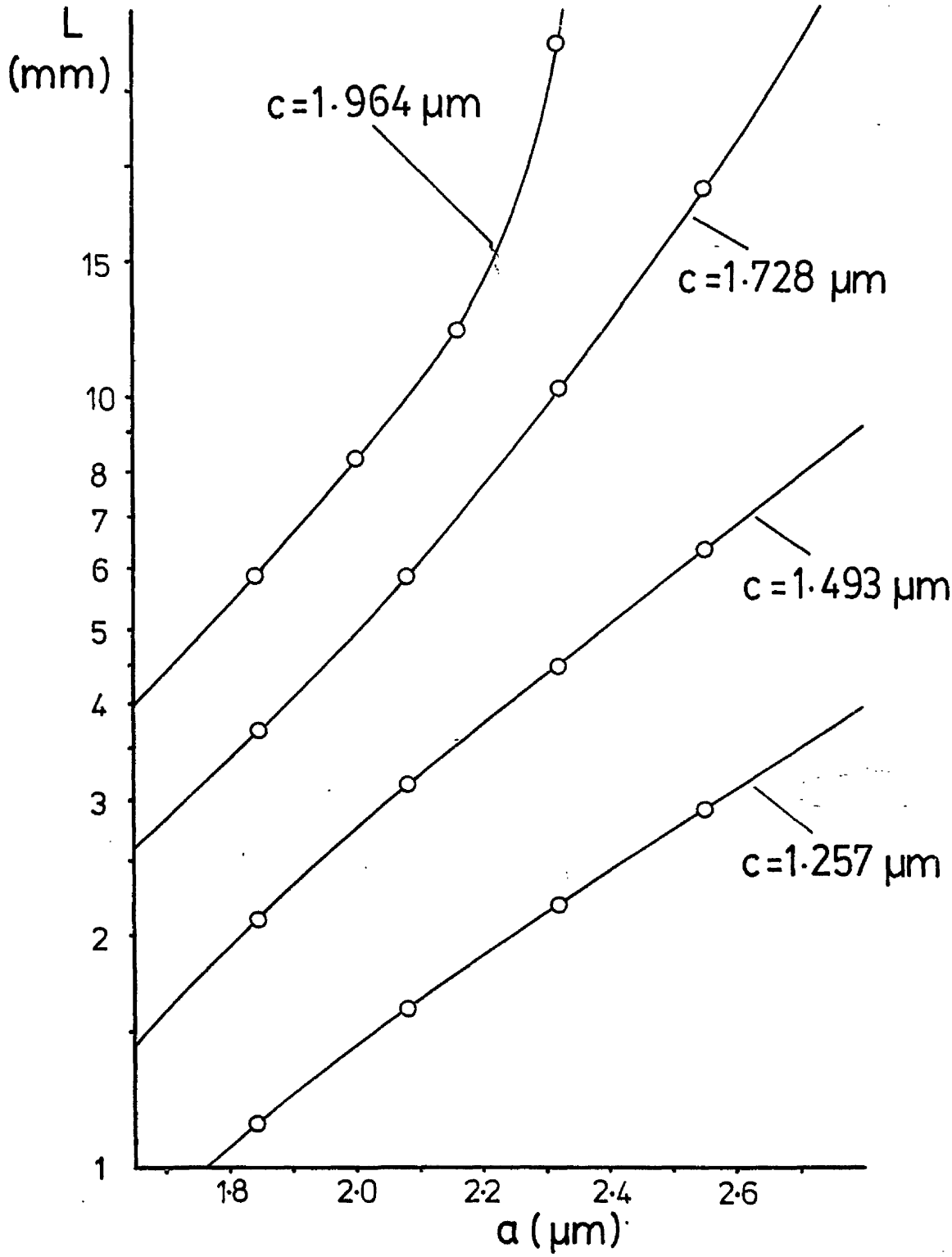


Fig. 9.4 Computed transfer length [L] as a function of directional coupler dimensions: a [aperture width] and c [aperture separation] slab n_e=1·5253

Table 9.1 Comparison of measured and computed transfer lengths for experimental directional couplers.

Slab n_e = 1.5253

	Coupler Dimensions			I_2/I_1 (dB)	Transfer Length L (mm)			% Error L_m w.r.t. L_c
	a (μm)	c (μm)	s (mm)		Computed (exact)	Measured (probable)	Measured (alternative)	
A_1	2.34	1.53	4	+ 7.1	5.25	5.44	3.16	+ 3.6 %
B_1	2.27	1.53	2	- 1.83	4.65	4.61	1.28	- 0.86%
C_1	2.36	1.51	1	- 7.95	5.1	4.13	0.57	- 19 %
B_2	2.30	1.47	2	- 3.43	4.05	5.30	1.23	+ 31 %
C_2	2.44	1.33	1	- 6.14	3.15	3.43	0.58	+ 9 %
Tol.	± 0.05	± 0.05	-	± 0.5	± 15%	-	-	

Key a = aperture width

c = aperture separation

s = interaction length

L_m (L_c) = measured (computed) transfer length

Further Results

In order to extend the sample and to check the comparison at other diffusion times, further couplers were made and measured. The two coupler sets on each slide were made using different exposure times and it is found that most of the dimensions fall into one or other of two main groups. The nearest computationally possible dimensions to the average values of each group are:

(i) $a = 2.435$, $c = 1.414$; (ii) $a = 2.278$, $c = 1.65$ microns

Twenty hours of computer (c.p.u.) time are invested in the curves of Fig. 9.4. Therefore, rather than repeat this for each different value of the slab effective index, it was proposed that only two couplers, with the above dimensions, be considered for computing the variation of L with diffusion time. The resulting graph, which is shown in Fig. 9.5, may then be used in conjunction with Fig. 9.4 to estimate the theoretical transfer length at other diffusion times. Typical results are presented in Tables 9.2 and 9.3.

In general, these measured results compare less favourably with the theory than do those of Table 9.1. There is a strong tendency for the measured transfer length to be smaller than the computed, the average error being -13.6% in Table 9.3 and -19% in Table 9.2. The latter is increased by the coupler A_1 whose error of -56% is probably partially due to imperfections in the guides - especially likely when the interaction length is as long as 4 mm.

With the deeper diffusion (Table 9.2, slab $n_e = 1.5247$) the measured values of L are unambiguous. Likewise, with the couplers C_1 and C_2 ($s = 1$ mm) of the shallow diffusions (e.g. Table 9.3, slab $n_e = 1.5185$) there is no real choice, and a substantial negative error must be accepted. Other slides from the same batch of devices confirm this.

Table 9.2 Results for directional couplers

Slab n_e = 1.5247

	Coupler Dimensions			I_2/I_1 (dB)	Transfer Length L (mm)			% Error L_m w.r.t. L_c
	a (μm)	c (μm)	s (mm)		Computed (approx)	Measured (probable)	Measured (alternative)	
A_1	2.24	1.59	4	- 7.31	5.3	2.3	15.5	- 56 %
B_1	1.92	1.84	2	- 3.73	5.1	5.44	1.22	+ 6.7 %
C_1	2.12	1.71	1	- 8.19	5.5	4.23	0.57	- 23 %
A_2	2.38	1.4	4	- 0.95	3.5	2.61	1.62	- 25 %
B_2	2.36	1.39	2	+ 2.81	3.2	3.32	1.43	+ 4 %
Tol.	± 0.05	± 0.05	-	± 0.5	± 15%	-	-	

Key a = aperture width

c = aperture separation

s = interaction length

L_m (L_c) = measured (computed) transfer length

Table 9.3 Results for directional couplers

Slab n_e = 1.5185

	Coupler Dimensions			I_2/I_1 (dB)	Transfer Length L (mm)			% Error L_m w.r.t.L_c
	a (μm)	c (μm)	s (mm)		Computed (approx)	Measured (probable)	Measured (alternative)	
A_1	2.34	1.64	4	- 0.44	2.7	2.64	1.61	- 2.2 %
B_1	2.18	1.76	2	+ 7.43	2.8	2.68	1.59	- 4.3 %
C_1	2.37	1.61	1	+ 2.64	2.6	1.68	0.71	- 35 %
A_2	2.50	1.4	4	+ 1.71	1.8	1.56	2.78	- 13 %
B_2	2.35	1.45	2	+ 7.15	1.8	1.58	2.7	- 12 %
C_2	2.54	1.36	1	+ 6.32	1.65	1.4	0.77	- 15 %
Tol.	± 0.05	± 0.05	-	± 0.4	± 18%	-	-	

Key a = aperture width

c = aperture separation

s = interaction length

L_m (L_c) = measured (computed) transfer length

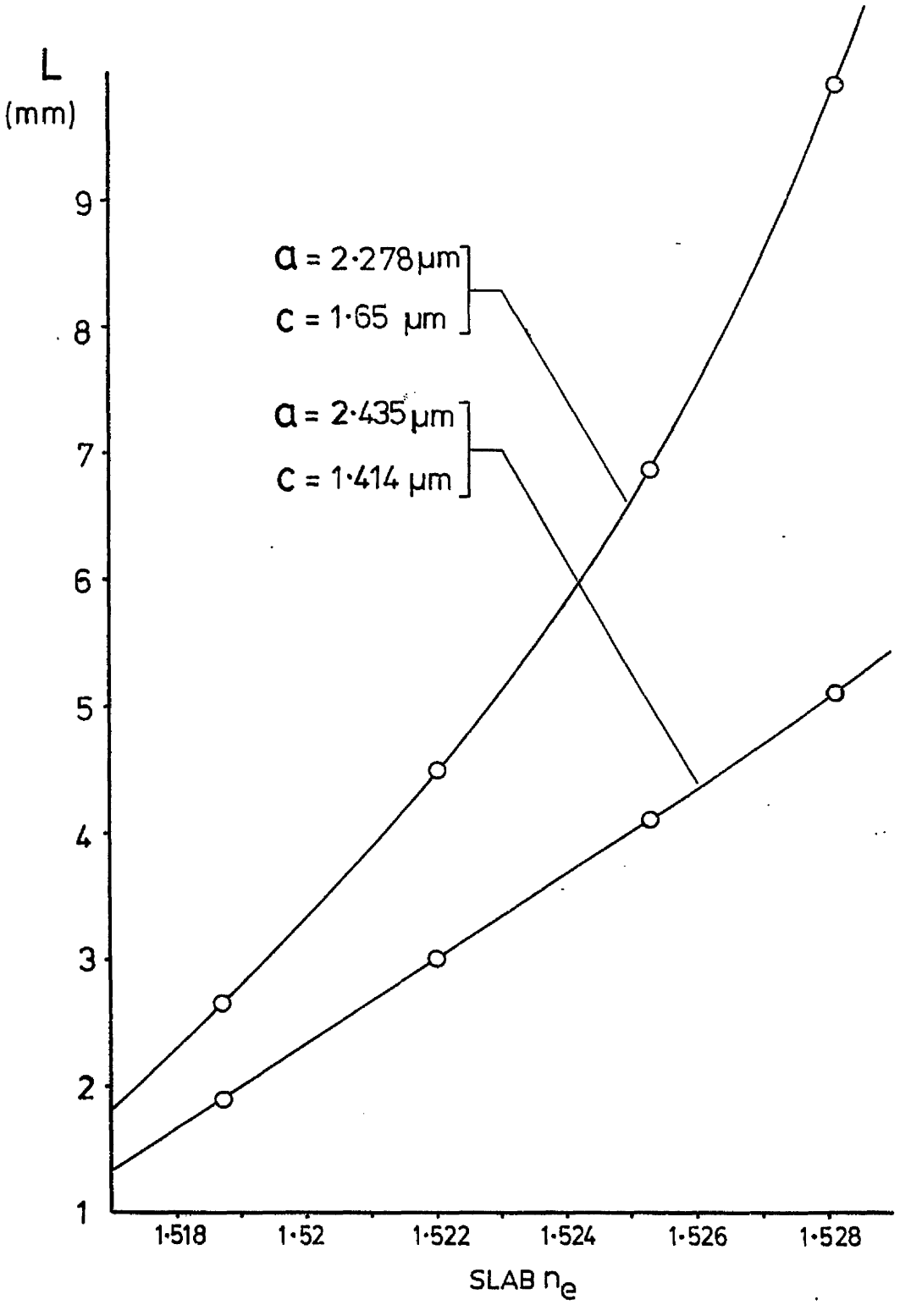


Fig. 9·5 Computed transfer length [L] as a function of the slab n_e for two sets of coupler dimensions.

Discussion of Errors

The consistent, negative error shown by the above batch of couplers is probably due to the anodised masks being less than perfect. Any permeability of the mask to silver ions would raise the refractive index of the substrate and thus reduce L.

The fundamental modes of some of these couplers were measured and were found to be not inconsistent with the "improved, anodised" characteristic (Fig. 8.3), i.e. the measured n_e is slightly greater than the theoretical. It is possible that the sample of Table 9.1, whose average error is +4.7%, had a particularly good anodised mask.

The dimensional tolerances quoted in the tables are deliberately generous to allow for variations along the coupled length; the tolerance on the computed value of L is based on these (a + 0.05 and c + 0.05 gives one extreme). It is not possible to give a standard tolerance for the measured transfer length based on the suggested tolerance on the intensity ratio (which is admittedly speculative). Each alternative value of L_m (measured L) has a different tolerance; when m is large (see 9.2.2) KS is dominated by $m\pi$.

As has already been suggested, a scattering spot or defect on the coupler could seriously affect the apparent transfer length, as could a disturbance of the input coupling conditions between the intensity readings. However, the x, z and θ_i (the input angle) controls were always adjusted for maximum output before any reading was taken.

Fig. 9.6 shows photographs of various coupler output-arms (50 x magnification). It is instructive to compare the observed brightness differences with the measured intensity ratios.

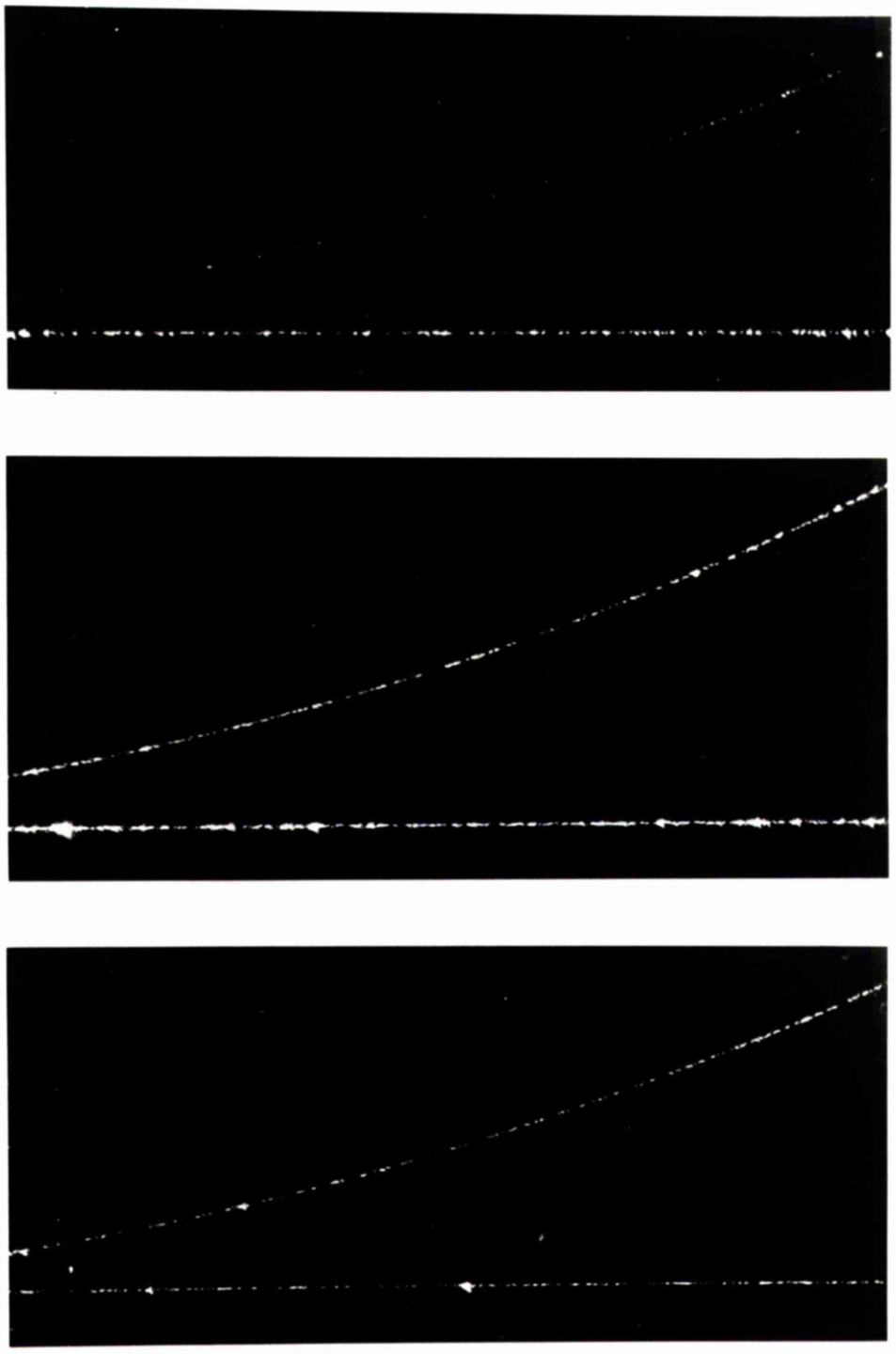
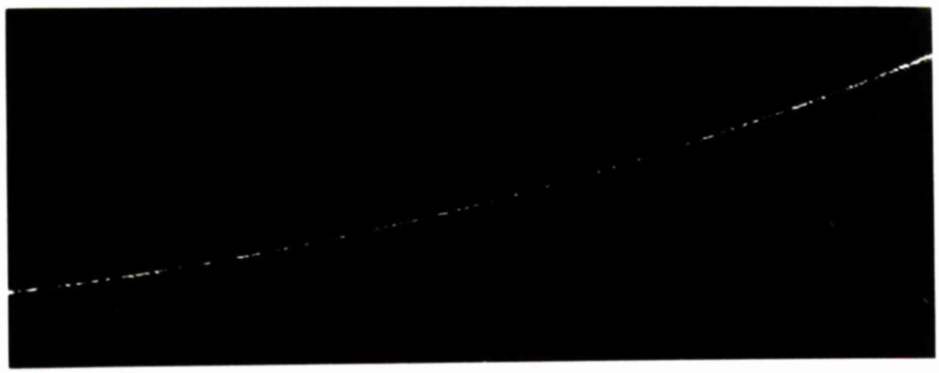


Fig. 9·6 Various directional coupler output arms. The measured intensity ratio is indicated in dB.

(continued overleaf)

+2

+5



+12·5

(Fig. 9·6 continued)

Conclusions

It has been shown that DIF and WAVE (i.e. the mathematical model) are capable of accurately predicting the transfer lengths of ion-exchanged directional couplers. Such couplers have transfer lengths which increase with width, guide separation and diffusion depth. However when the transfer length is small, it is less sensitive to these parameters than when large.

9.3 Curved Stripe Waveguides

There is, at present, no entirely satisfactory way of measuring the constant-curvature loss of a curved, stripe waveguide. The most obvious method is to use a 90° bend (or two, in an 'S' curve) and measure the guided intensity before, and after, the curved section using an output prism. The main difficulty with this technique is one of accurately positioning the prism so as not to include a variable length of (variably lossy) straight guide. Also this measurement would include the discrete loss which occurs at each change of curvature but is not relevant to a circular ring-resonator.

It is possible to estimate attenuation by using a photo-sensitive probe to detect the surface-scatter; however this is probably too clumsy for use with bends of only about 100 μm radius. A similar idea is to photograph the surface scatter from the bend (using a microscope) and then measure the image density on the film.

The waveguide-bend pattern was designed with this technique of "photo-densitometry" in mind (see Fig. 9.7). The 180° ("hairpin") bends were designed to provide the maximum length of continuous curve but, of course, prevent the use of an output prism. The overall pattern is very

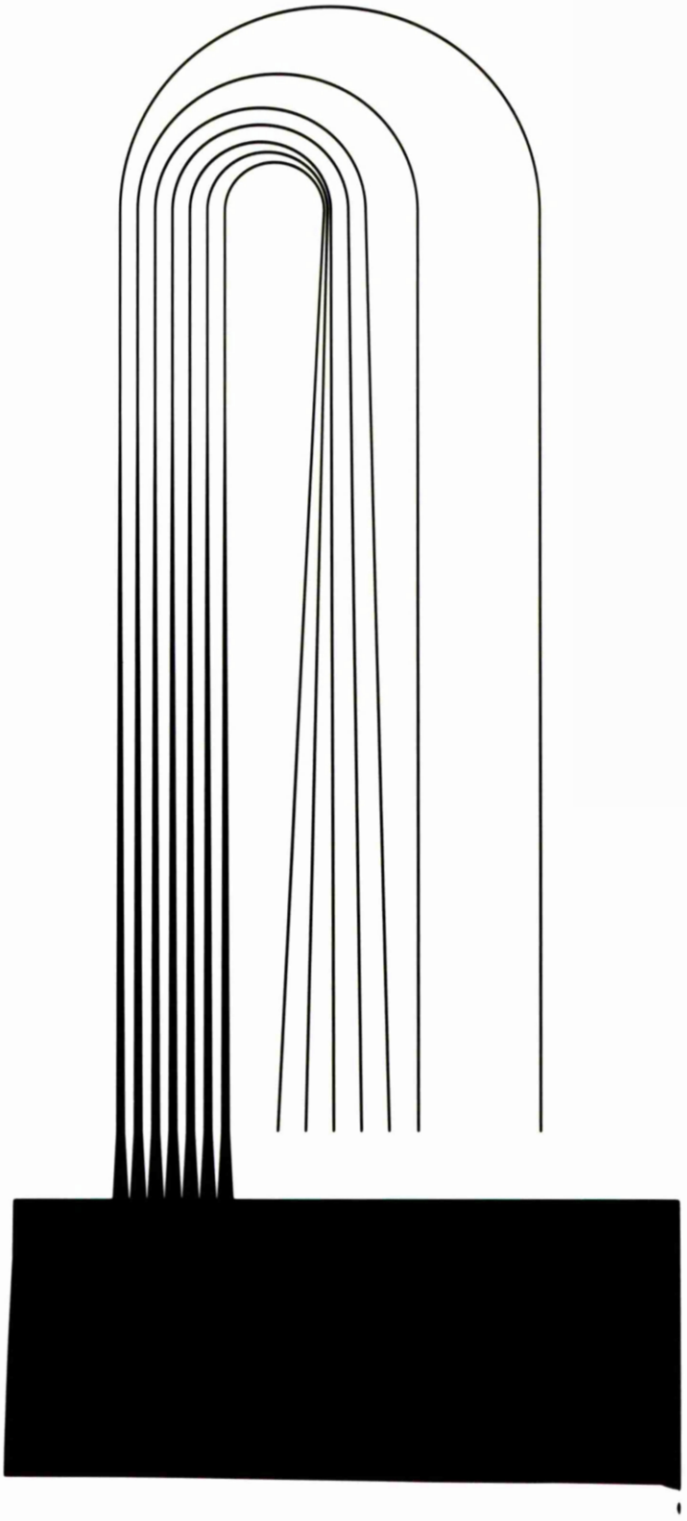


Fig. 9.7 The curved waveguide shadow-mask

small, being little more than 2 mm in length, as was essential for accurate cutting of the 2 μm waveguides; the low aspect-ratio allows a very large photo-reduction ratio to be used.

Light is guided into the bends via short, straight sections (sufficient to allow width measurement) from tapers. The small size of the pattern makes it essential that the light be first coupled into a slab region some distance from the taper entrances and focussed on to these.

Seven radii are available: 300, 200, 150, 125, 100, 85 and 70 microns, which may be excited in turn by vertical translation or simultaneously by adjusting the focus of the lens to broaden the input beam.

As it happened, there was insufficient time to investigate the photo-densitometry technique. However, sufficiently accurate results were gained by much simpler means. To introduce this, some approximate theory in addition to that contained in Chapter 3, is necessary.

As the radius of a ring-resonator is increased the bend-loss decreases (see section 3.4) and the straight-guide loss (per circuit) increases. It is a fair assumption that the curvature attenuation is approximately proportional to that fraction of the guided energy which is beyond the critical radius (ρ_c). If the field amplitude varies exponentially, then this energy will be proportional to the square of the amplitude at ρ_c. Therefore, the curvature ring-loss (Γ_c dB/circuit) is:

$$\Gamma_c = \frac{K}{\xi} \exp(-2\xi c) \qquad (9.3.1)$$

where $$c = \rho_c - R = R\left[\frac{n_e}{n_s} - 1\right] .$$

and where ξ is the evanescent decay constant

R is the radius of curvature of the outer guide boundary

n_s is the substrate refractive index

K is a constant.

Thus Γ_c varies (approximately) exponentially with R whereas the straight-guide ring-loss (Γ_L) varies linearly with R:

$$\Gamma_L = \frac{2\pi R}{10^4} \alpha_L \quad \text{dB/circuit}$$

where R is in microns and α_L in dB/cm.

The total ring-loss ($\Gamma = \Gamma_c + \Gamma_L$), therefore, goes through a minimum value for some radius. It is found that Γ_c varies with R much more rapidly than does Γ_L, therefore at this optimum radius, Γ is dominated by Γ_L and the precise value of Γ_c is umimportant.

The Experimental Method

Several slides were prepared, each containing six copies of the bend pattern of Fig. 9.7 and one double-sized copy (i.e. 4 μm wide guides with a maximum radius of 600 μm). These were diffused for various times (t) between 5 and 25 minutes. Due to the use of MF 312 developer some of the guides were narrower than 2 μm; these were given the longer diffusion times so that as many as possible of the guides would be single-mode .

Assessment of the curvature losses was by visual observation, with some aid from photography. The directional coupler photographs (Fig. 9.6) confirm that an intensity difference of 2 dB can be detected by the eye. Using direct observation this difference is easier to detect if the input coupling is adjusted to reduce the guided intensity almost to zero; thus small intensity differences become more noticeable.

The aim of the technique is to note a cut-off radius (R_c) for each slide. This is the largest radius for which some diminution of brightness around a semi-circular curve is detectable using the eye and a 50 x magnification microscope. R_c is taken to be that radius for which 3 dB is lost per semi-circuit, therefore $\Gamma(R_c) \simeq 6$ dB. This is a generous estimate and $\Gamma(R_c)$ may well be rather lower than this.

The method may be explained by reference to the photographs of Fig. 9.8 showing light being guided around the curves (from bottom to top). It was decided, in this case, that no loss could be detected in the three largest curves (300, 200, 150 μm) but that some was detectable in the fourth (125 μm). R_c was estimated at 135 μm, using the degree of attenuation in the 125 μm curve to interpolate. All six curve sets were examined and the most defect-free curve used for each estimate. Brightness changes at the junction of the straight and curved sections were ignored. Finally, the slab effective index is measured and used to estimate the actual stripe n_e using the measured guide widths and the graphs of Fig. 8.3 (the pattern is too small to measure n_e directly). $1/R_c$ is then plotted against n_e.

The relationship between $1/R_c$ and $(n_e - n_s)$ should be almost linear if the assumptions leading to 9.3.1 are valid. Rearranging 9.3.1 we obtain:

$$\frac{1}{R_c} = \frac{2\xi}{n_s} \cdot \frac{1}{\ln\left[\frac{K}{\xi \Gamma_c(R_c)}\right]} \cdot [n_e - n_s]$$

In this experiment the guides are sufficiently wide that the small variations in width will not much affect ξ. Also, since ξ describes the lateral evanescent field it may be expected to be relatively insensitive to depth, except near cut-off where $\xi \rightarrow 0$. $1/R_c$ (above) is insensitive to such variation in ξ (unlike the transfer length of a directional coupler) because ξ, in the above relation, is never exponentised. Thus, the graph of $1/R_c$ against $[n_e - n_s]$, where n_e is varied by varying diffusion depth, should be nearly linear with some curvature near cut-off.

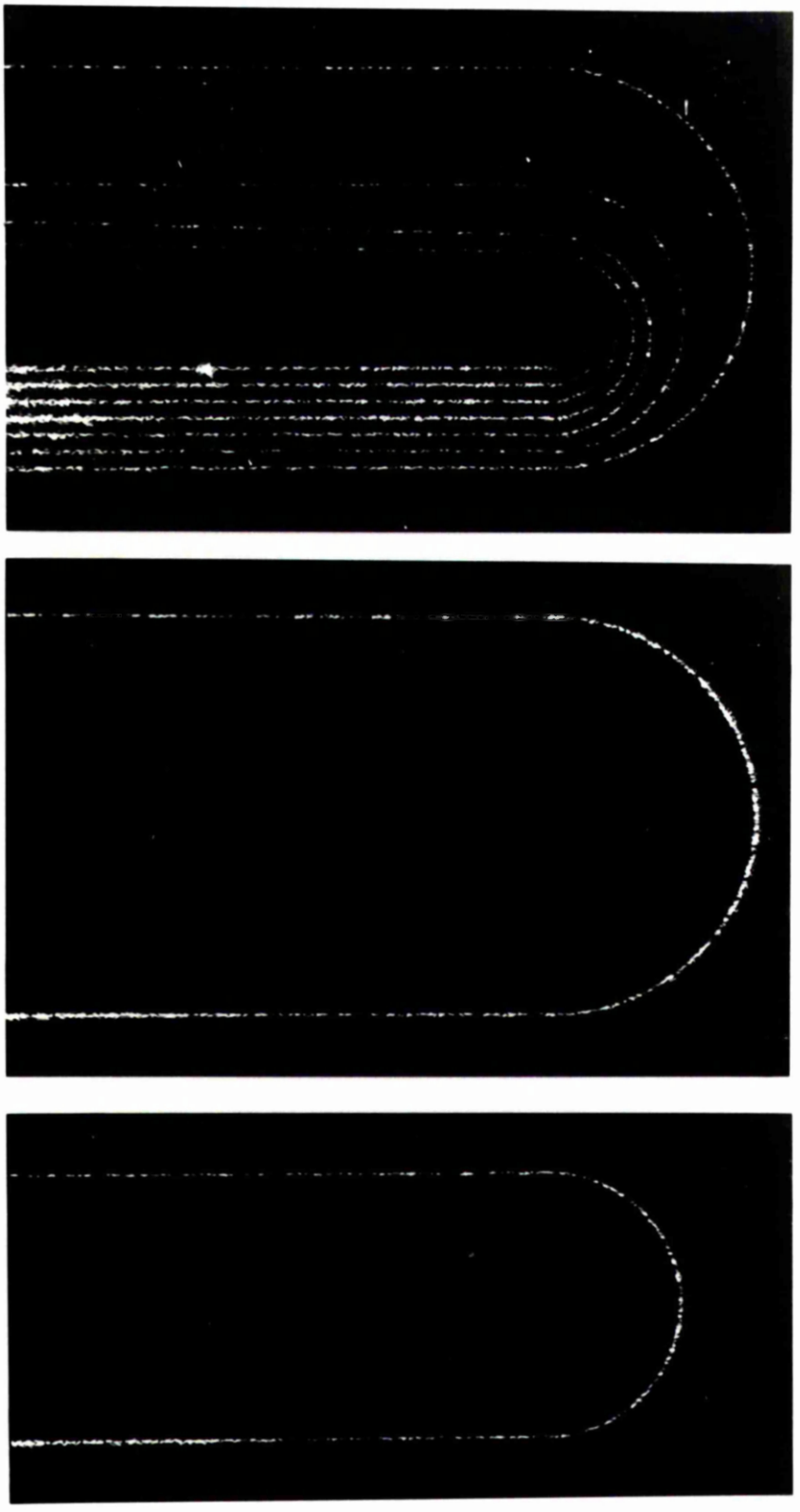
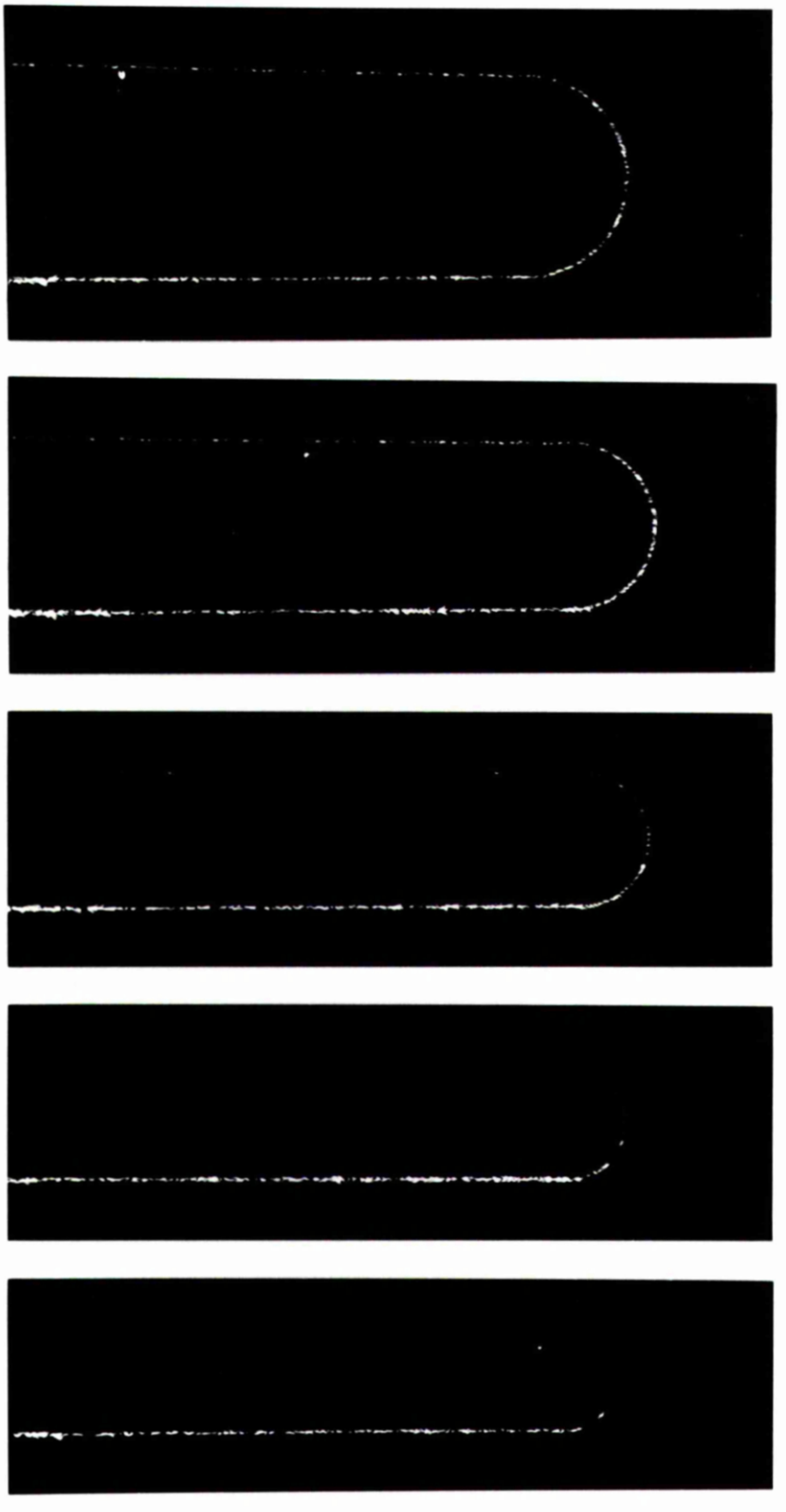


Fig. 9.8 Surface scatter from curved waveguides: guide width ≃ 2 μm ; slab n_e = 1·5327

(continued overleaf)



(Fig. 9·8 continued)

Results

An experimental graph of this type is presented in Fig. 9.9 and shows the expected variation. One feature, however, requires explanation. The highest point was from an early batch of devices and probably conforms to the "improved anodised" characteristic (i.e. good quality masking). The remaining points are from devices made at the same time as the linear attenuation samples (see section 9.1) and may have the same faults (i.e. a somewhat permeable diffusion mask resulting in a reduced lateral index difference, reduced mode separations, and n_e close to the slab n_e). Because such imperfect anodised masks tend to produce a stripe n_e which lies between the theoretical and the slab values, these values are used to mark the extrema of the horizontal error bars. When drawing the line, more weight was given to the data from the earlier batch than to these.

Post-Baked Curved Waveguides

The same samples as above were post-baked, the observations being repeated after post-bake times of t and 3t (where t is the initial diffusion time). In Fig. 9.10, $1/R_c$ is plotted against n_e (n_e after post-baking) for the two sets of post-baked results. For the sake of comparison the line from Fig. 9.9 is also included.

As may be expected, post-baking increases the bending-loss by reducing n_e and also by reducing the evanescent decay constant (ξ). Fig. 9.10 shows only the latter effect. The shallowest diffusion shows an anomalous reduction in the cut-off radius after a 3t post-bake, due to the substantial increase in n_e. However, at the pivotal diffusion time t_p (no change in n_e) R_c increases. This is probably, at least partially, due to the lateral diffusion profile becoming more graded and less rectangular.

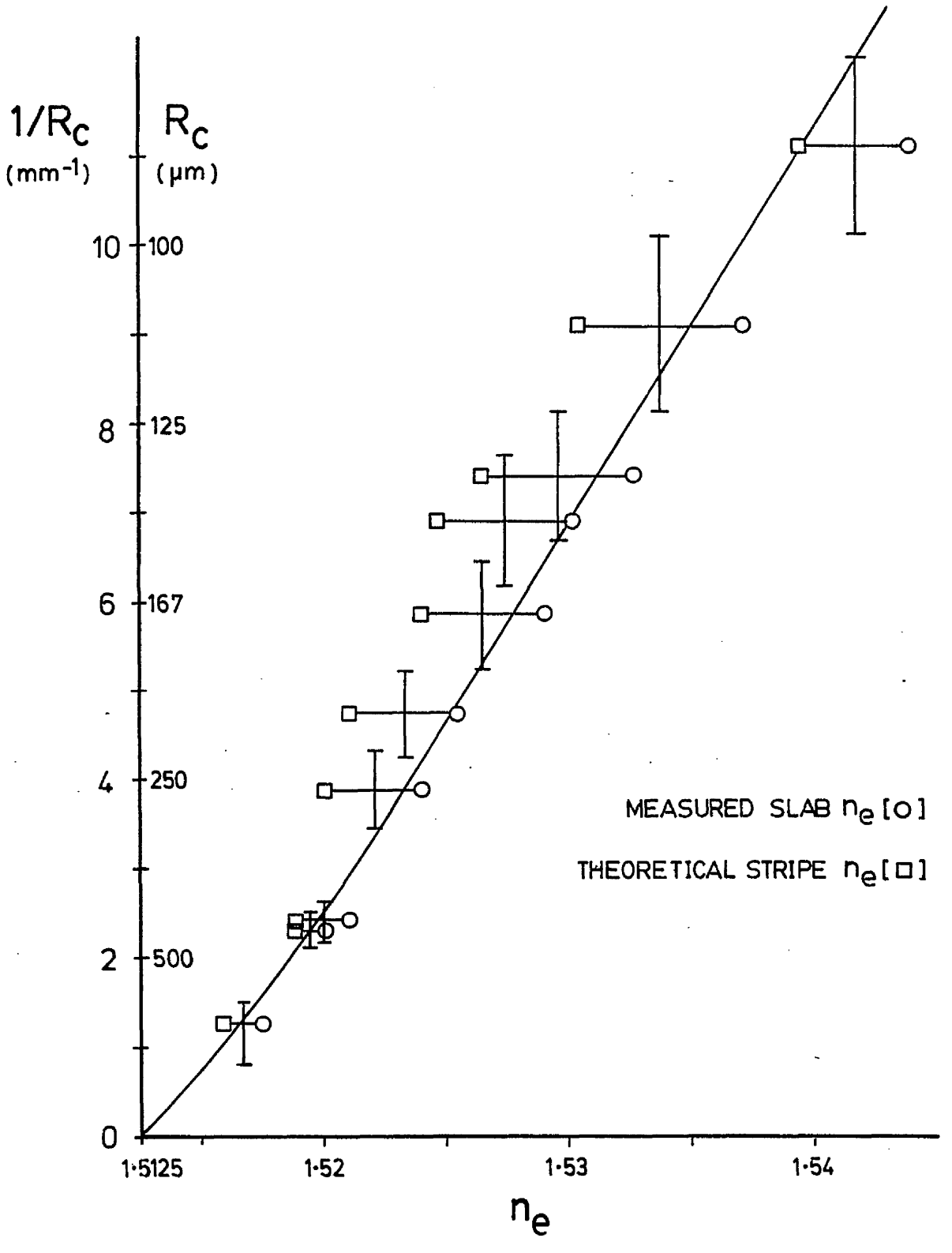


Fig. 9·9 Estimated cut-off radius : reciprocal [$1/R_c$] plotted vs. n_e. [curved stripe waveguides ; varying diffusion times (t)]

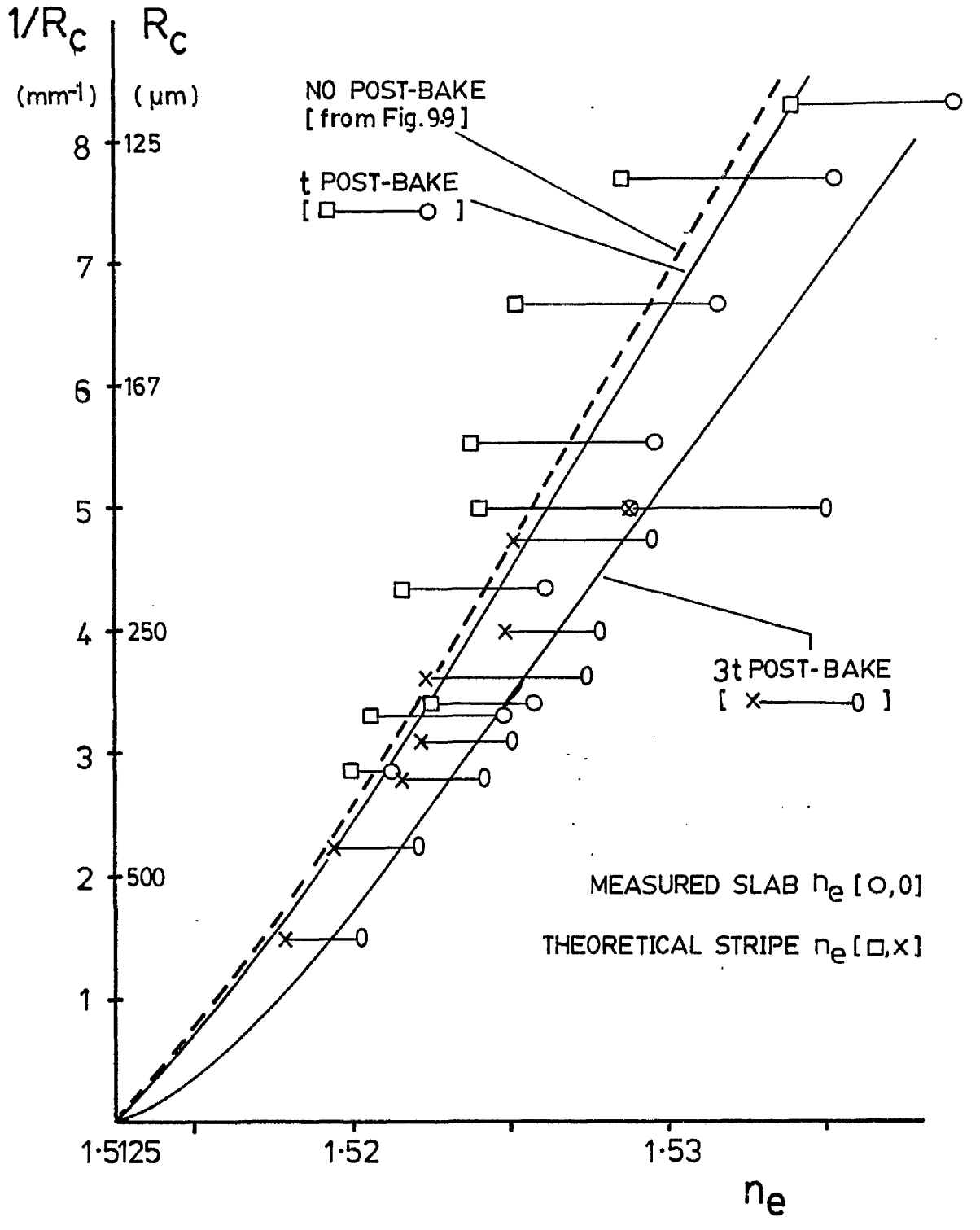


Fig. 9·10 Reciprocal cut-off radius vs. n_e for post-baked, curved waveguides.

[vertical error-bars omitted for clarity]

It is known (Heiblum and Harris[31] show this) that diffused boundaries produce evanescent fields which spread further than those of step boundaries.

9.4 Ring-Resonator Design

On the basis of the data presented in sections 9.1-9.3 it is possible to suggest some design parameters for a practical ring-resonator filter.

(i) Guide Dimensions. As has already been suggested, the guides should be as narrow as possible to obtain strong directional coupling, and to allow single-mode guidance to be consistent with a large value of n_e (required for the sake of low bending-loss). The narrowest practical width is 2 microns, giving single-mode guidance after 10 minutes diffusion (maximum slab $n_e = 1.528$, stripe $n_e = 1.523$).

(ii) Post-Baking. By a comparison of Fig. 9.10 with Fig. 9.2, it is seen that a short post-bake achieves the maximum improvement in linear attenuation for minimum increase in bending loss. Thus it is proposed that the device be both diffused and post-baked for approximately 10 minutes.

(iii) The Ring Radius. The optimum radius is that for which the total ring-loss ($\Gamma = \Gamma_L + \Gamma_c$) is minimised. If the variation of Γ_c with R is given by 9.3.1 then:

$$\Gamma = \frac{2\pi\alpha_L}{10^4} R + K_1 \exp\left(-2\xi\left[\frac{n_e}{n_s} - 1\right]R\right) \qquad (9.4.1)$$

The optimum radius, R_0, is obtained by differentiating the above with respect to R and setting to zero thus:

$$R_0 = \frac{1}{2\xi\left[\frac{n_e}{n_s} - 1\right]} \cdot \ln\left[\frac{K_1\xi\left[\frac{n_e}{n_s} - 1\right].10^4}{\pi\alpha_L}\right] \qquad (9.4.2)$$

where $K_1 = K/\xi$ (see 9.3.1)

and ξ is assumed to be constant with R (the effect of the curvature is small for $R > R_c$ - i.e. low loss).

The value of ξ is obtained from a computer-plot of the lateral field amplitude (using 12 x-direction basis functions) for a 2 μm diffusion-aperture width. The initial diffusion is followed by a post-bake of 1 t_n giving a final stripe n_e of 1.5211 - a little lower than the maximum for single-mode guidance. A series of points on the evanescent tail of the curve (see Fig. 9.11), each in ratio 1/e with its higher neighbour, are separated horizontally by $1/\xi$. Thus Fig. 9.11 gives an average value for ξ of 2.56 μm^{-1}. In order that the design be not over-optimistic (the curvature tends to reduce ξ - see section 3.4) we take $\xi = 2.2\ \mu m^{-1}$ (say).

The constant K_1 is fixed by noting that when $R = R_c$, $\Gamma \simeq 6$ dB. Allowing some margin of safety (to ensure that the guide is single-mode) we choose $n_e = 1.522$, and (from Fig. 9.10) $R_c = 300$ μm. Assuming $\alpha_L = 3$ dB/cm and $\Gamma_c \simeq 5.5$ dB then (from 9.4.1) $K_1 = 21{,}931$ dB.

Thus, from 9.4.2 and 9.4.1 we obtain the optimum ring-resonator design parameters:

$$R_0 = 459\ \mu m$$

$$\Gamma = 0.933\ \text{dB (per circuit)}$$

The resulting ring-resonator (with optimum coupling) has a theoretical acceptance ratio (see section 3.2, equation 3.2.10) of 19.4 dB ($\gamma = 0.898$) and a Q-factor of 1.014×10^5 (equation 3.2.8). The resonant-frequency spacing is 68.346 GHz and the half-power bandwidth, δf, is 4.56 GHz.

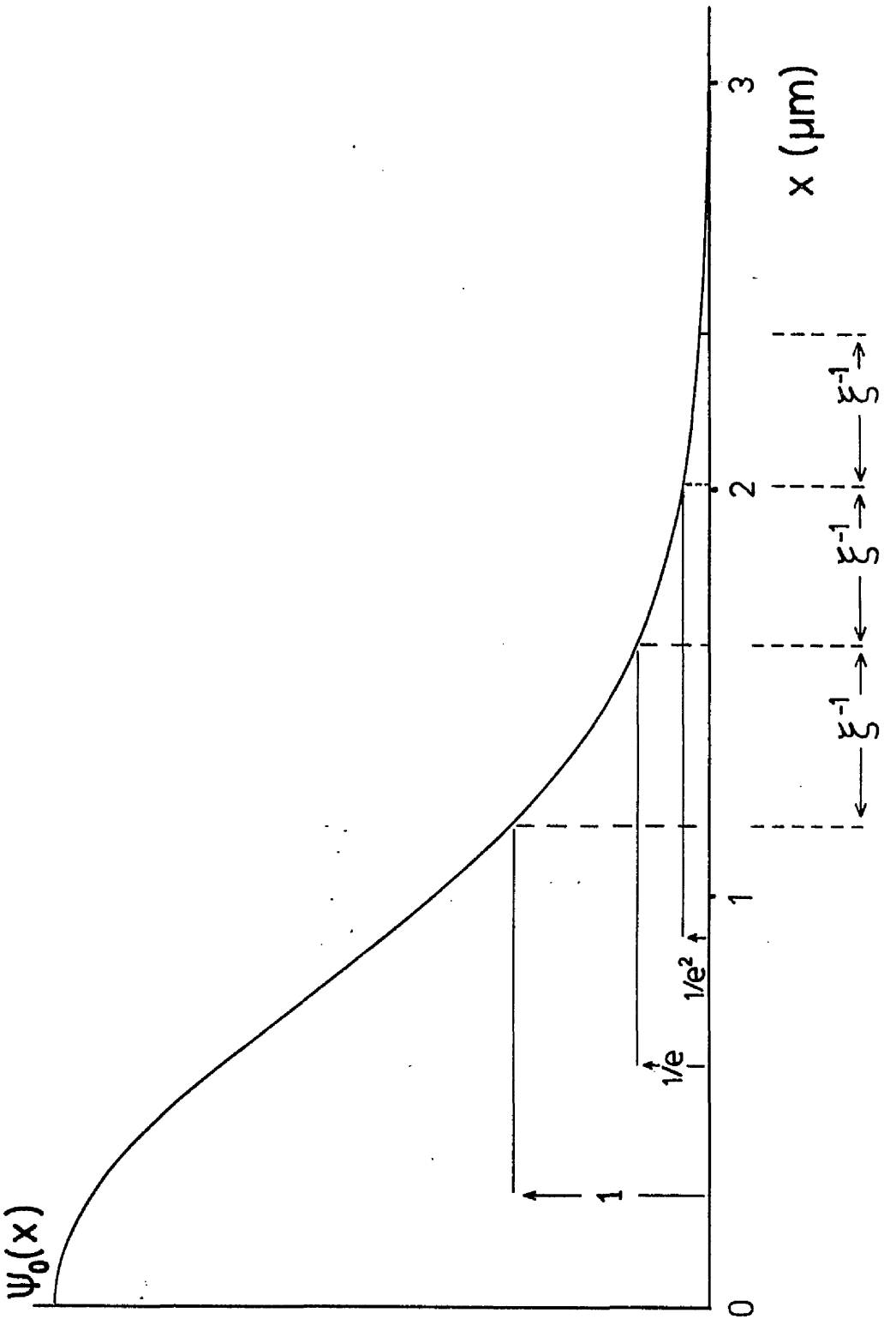


Fig. 9.11 Computed lateral field-amplitude distribution, indicating the derivation of ξ – the evanescent decay constant.

(iv) The Input Coupler. Since $\gamma = 0.898$, an interaction length of $s = 0.29\ L$ (where L is the transfer length) is required for optimum coupling (see 3.2.5). If the input guide follows the curve of the ring for a full quarter-circle (the maximum practical amount, if output coupling is also required) then a transfer length of 2.49 mm is needed.

To judge the practicality of this, a single, post-baked coupler was computed, with aperture width (a) of 2.0 μm and aperture separation (c) of 1.493 μm. The equivalent diffusion time was 9.68 minutes (assuming the value of D_a found in Chapter 7) giving an initial slab effective index of 1.5284.

The computed transfer length was 3.44 mm after the initial diffusion, and 1.47 mm after a post-bake of 1t. Thus, it should be possible to relax the fabrication requirements, making c rather larger.

(v) Output Coupling. It is not proposed that output coupling be considered until working ring-resonators have been demonstrated. Output coupling will degrade the resonator performance and may be impractical until the ring-loss, Γ (which is dominated by the straight-guide loss, α_L) has been reduced by technological improvements.

In conclusion it must be stated that the design possibilities have not been exhausted by this treatment. For example, an increased post-baking time would reduce the linear attenuation but increase the optimum radius. The effect on the overall ring-loss cannot be stated without performing the necessary calculations, however it would undoubtedly relax the input coupling requirements, both by increasing the coupling coefficient, and increasing the available interaction length.

CHAPTER 10

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

It has been shown that silver/sodium ion-exchange in soda-lime glass is capable of producing good quality stripe (2-D) optical waveguides. A short post-diffusion bake is found to reduce the attenuation to less than 3 dB/cm for a 2 μm wide waveguide. Low-loss propagation around waveguide bends (180°) with radii between 100 and 300 μm has been demonstrated; the cut-off radius depends upon the diffusion depth.

In order to achieve these results, it is necessary to use an electrically insulating diffusion mask; anodised aluminium was used in this work.

In retrospect, it seems as though the fabrication of couplers, bends and stripes for attenuation measurements was a little premature since the anodisation process was not yet fully under control. Nevertheless, however imperfect the results, the information so gained has proved very valuable and has allowed parameters for a practical ring-resonator to be suggested.

A working resonator has not yet been made. Before this is attempted, it is proposed that the anodisation process be more thoroughly investigated to enable waveguides whose propagation characteristics match the computed theory to be reliably made. It is proposed that the tartaric acid/ammonium tartarate anodisation system, which produces a dense, impervious oxide layer, be investigated.

The substantial agreement between the measured and computed transfer lengths for directional couplers is encouraging, however the transfer length required for a practical ring-resonator is very short and it may be difficult to make such a coupler accurately. This limitation only

applies at short wavelengths such as that used in this work (λ = 0.6328 μm). Practical optical communication systems will operate in the near infra-red ($\lambda \approx 1.3$ μm). At these wavelengths, most of the dimensions discussed in section 9.4 may be approximately doubled, thus greatly relaxing the fabrication tolerances.

Clearly, the mathematical model (Chapter 5), though quite successful, has several deficiencies, the most important of which is the discrepancy between the real and the computed depth diffusion-profile. As discussed in Chapter 7, this seems to be due to additional non-linearity which is not accounted for in the differential equation. It may be of benefit to try some different forms for the inter-diffusion coefficient, to achieve a better match empirically.

APPENDICES

APPENDIX 1

The Scalar Wave Equation

The wave equation is derived from Maxwell's curl equations:

$$\text{curl } \underline{H} = \frac{\partial(\varepsilon \underline{E})}{\partial t} \quad ; \quad \text{curl } \underline{E} = -\mu_0 \frac{\partial H}{\partial t}$$

It is (reasonably) assumed that $\mu = \mu_0$ and $\rho_v = 0$.

Taking curl $\underline{E}$ and eliminating the resulting curl $\underline{H}$ term we obtain:

$$\text{grad div } \underline{E} - \nabla^2 \underline{E} = -\mu_0 \varepsilon \frac{\partial^2 \underline{E}}{\partial t^2} \qquad \text{(A1.1)}$$

Since $\text{div } \varepsilon \underline{E} = 0$ we may say:

$$\text{div } \underline{E} = - \frac{\underline{E}.\text{grad } \varepsilon}{\varepsilon}$$

therefore $\text{div } \underline{E} = \underline{E}.\text{grad } (\ln \varepsilon)$

The vector wave-equation A1.1 may be reduced to three scalar wave-equations, one for each component of $\underline{E}$. For E_x:

$$\frac{\partial}{\partial x} [\underline{E}.\text{grad}(\ln \varepsilon)] - \nabla^2 E_x = -\mu_0 \varepsilon \frac{\partial^2 E_x}{\partial t^2} \qquad \text{(A1.2)}$$

Approximation 1

We consider only quasi-T.E. modes in which E_x is the dominant field component.

If $E_x >> E_y$ then $\underline{E}.\text{grad}(\ln \varepsilon)$ becomes:

$$E_x \frac{\partial}{\partial x} (\ln \varepsilon)$$

The solution will be of form:

$$E_x(x,y,z,t) = E_x(x,y) e^{j(\omega t - \beta z)}$$

Thus A1.2 becomes:

$$\frac{\partial E_x}{\partial x} \cdot \frac{\partial}{\partial x} (\ln \varepsilon) + E_x \frac{\partial^2}{\partial x^2} (\ln \varepsilon) - \nabla^2 E_x + \beta^2 E_x = k^2 E_x \qquad \text{(A1.3)}$$

where $k^2 = \omega^2 \mu_0 \varepsilon$ and E_x refers only to the amplitude function $E_x(x,y)$

Approximation 2

It is usual to neglect the terms involving ln ε in order to obtain a solution. This can be justified by reference to such a solution for E_x, obtained using the computer model of Chapter 5. We consider the 2-D diffusion profile of Fig. 5.4 (a) and the electric field amplitude plot of Fig. 5.13 (fundamental mode). Fig. 5.4 (a) may be considered to show refractive index contours [n(x,y)]; ε(x,y) is proportional to $n^2(x,y)$.

Clearly, $\frac{\partial}{\partial x} (\ln \varepsilon)$ is nearly zero over most of the guide but may be significant near the edges. Accordingly, we consider the magnitudes of all the terms of A1.3 near the guide edge. We take $E_{max} = 1$, x and y in microns and consider the refractive index to vary between 1.51 and 1.6.

The following values are obtained:

$$E_x \simeq 0.2 \quad ; \quad \frac{\partial}{\partial x} (\ln \varepsilon) \simeq 0.1 \quad ; \quad \frac{\partial E_x}{\partial x} \simeq 0.53$$

$$k^2 \simeq \beta \simeq 240 \quad ; \quad \frac{\partial^2}{\partial x^2} (\ln \varepsilon) \simeq 0.15 \quad ; \quad \frac{\partial^2 E_x}{\partial x^2} \simeq 0.56$$

Replacing each term of A1.3 by its numerical value, we obtain:

$$(0.053) + (0.03) - 0.56 - \frac{\partial^2 E_x}{\partial y^2} + (\sim 48) = (\sim 48)$$

$\frac{\partial^2 E_x}{\partial y^2}$ varies with y but is probably quite large near the maximum-intensity depth. Thus, it is seen that, even in the region where the first two terms are greatest, they are still of minor importance.

Conclusion

For Quasi-T.E. modes it is sufficiently accurate to solve the scalar wave equation:

$$\nabla^2_{xy} E_x + (k^2 - \beta^2) E_x = 0$$

or, if k is not constant,

$$(\nabla^2_{xy} + k^2) E_x = \beta^2 E_x \qquad \text{(A1.4)}$$

APPENDIX 2

Supplement to the Theory of Ion-Exchange

A2.1 The Composition of Glass

Ideally, the soda-lime glass should contain nothing but SiO_2 (84%), Na_2O (14%) and Al_2O_3 (2%); the alumina serving to prevent phase separation. The exact composition of "Fisher" glass is not available; but the composition of "Chance" glass, which is probably similar, is as follows:

SiO_2	72.1 %
Na_2O	14.0 %
Al_2O_3	1.7 %
K_2O	0.6 %
C_aO	7.3 %
M_gO	3.9 %
Fe_2O_3	0.1 %
SO_3	0.25%

Traces of iron and arsenic (for example) are harmful for silver/sodium ion-exchange since these elements cause silver precipitation.

Knowing the density of the glass (2.3 g/cm^2) and the molecular weight of Na_2O (= 62), the concentration of sodium ions in the above glass may be found:

$$C_K = 6.26 \times 10^{27} \text{ ions/m}^3$$

A2.2 The Charge Density Term

The instantaneous charge density, ρ, is given by Maxwells first equation:

$$\rho = \text{div } (\varepsilon \underline{E})$$

In order to estimate the magnitude of ρ we consider a one-dimensional diffusion and assume that ε varies slowly, thus:

$$\rho \simeq \varepsilon \frac{\partial E}{\partial y} \qquad \text{(A2.2.1)}$$

From the solution of section 4.4:

$$\frac{e}{kT} E = - \frac{\alpha}{1 - \alpha u} \frac{\partial u}{\partial y} \qquad \text{(A2.2.2)}$$

where u is the normalised silver concentration, $\frac{c_a}{c_k}$.

$$\frac{e}{kT} = 23.76 \text{ volts}^{-1} \text{ at } 215°C$$

and $$\varepsilon = \varepsilon_0 \varepsilon_r$$

where $$\varepsilon_0 = 8.85 \times 10^{-12} \text{ and } \varepsilon_r \simeq 2.4$$

Fig. A2.1 shows graphs of U(y) and $\frac{e}{kT} E$ (y) (normalised) for $\alpha = 0.7$. These were obtained using the computer solution of Chapter 5. We define the diffusion-depth to be that for which u = 0.5. Typically the diffusion depth is about 0.5 μm, therefore from Fig. A2.1, we obtain a surface grad u of -8.3×10^5; thus:

$$\frac{e}{kT} E \simeq 1.94 \times 10^6 \quad \text{at the surface}$$

$$E \simeq 8.18 \times 10^4 \quad \text{at the surface}$$

Having thus scaled the graph, we may find the maximum value of $\frac{\partial E}{\partial y}$:

$$\left(\frac{\partial E}{\partial y}\right)_{max} \simeq 1.04 \times 10^{11}$$

Therefore $\rho_{max} = 2.2 \text{ c/m}^3$ (from A2.2.1)

The density of vacancies required to cause such a charge density is ρ/e

$$\rho_{max}/e = 1.4 \times 10^{19}\ m^{-3}$$

This is insignificant, compared with C_K.

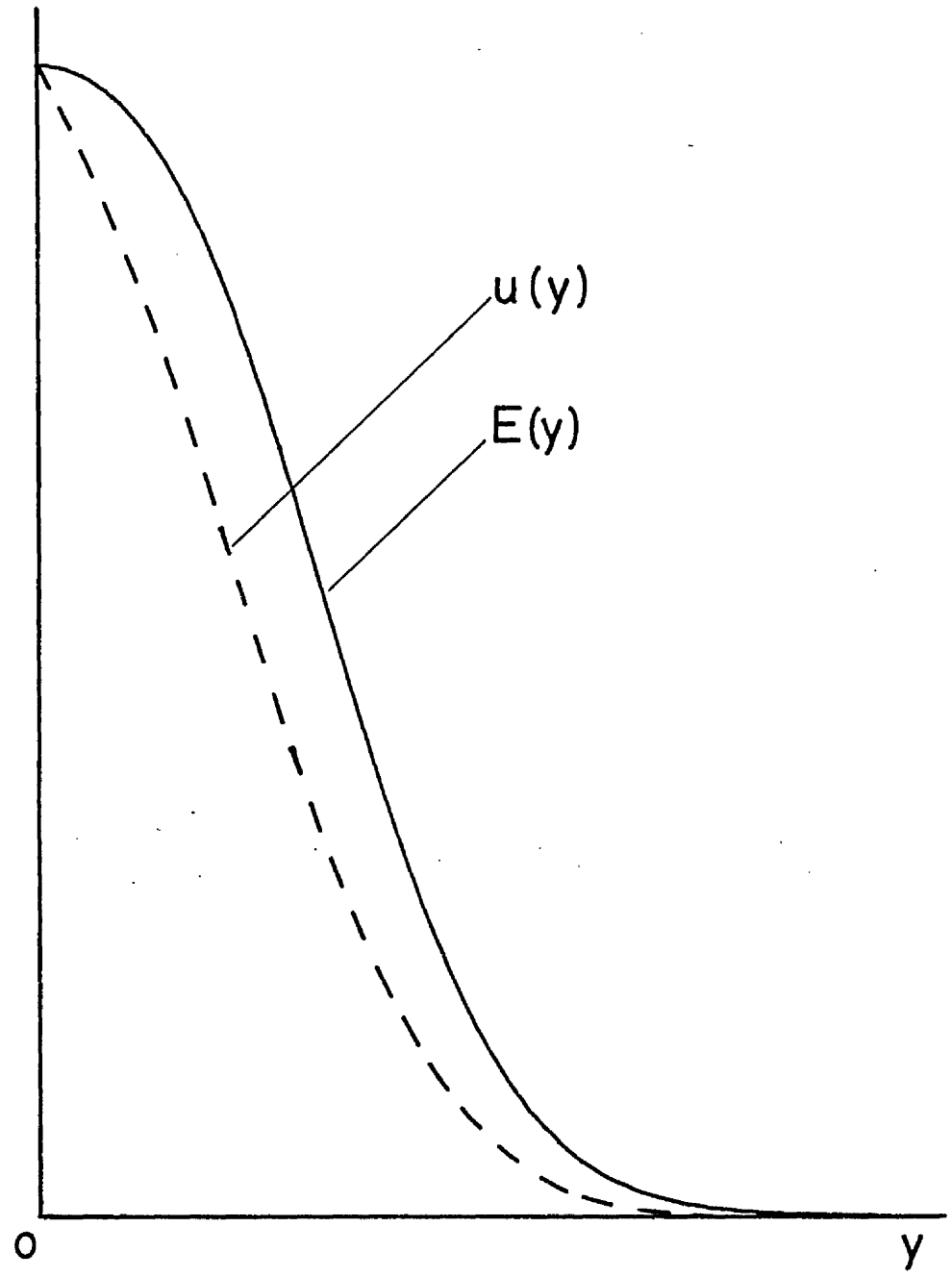


Fig. A2·1 Computed silver concentration [----] and electric field [——] during ion-exchange.
Both normalised.

APPENDIX 3

The Finite Difference Solution of the Diffusion Equation

A full discussion of finite difference techniques can be found in references (35) to (39).

A3.1 The Three Level Method

Three rectangular arrays are defined; U(x,y), LU(x,y) and NU(x,y) which contain concentration values at the three time-levels t, t - δt and t + δt respectively.

The time derivative, $\frac{\partial u}{\partial t}$ is replaced by its second-order correct finite difference approximation.

$$\frac{\partial u}{\partial t} \longrightarrow \frac{NU(x,y) - LU(x,y)}{2\delta t} \qquad \text{(A3.1.1)}$$

[In this algorithm, all gradients are centred about point U(x,y)].

The spatial derivatives are calculated using the central difference operators d_x and d_y where:

$$d_x(F) = [F(x + \tfrac{1}{2}\delta x) - F(x - \tfrac{1}{2}\delta x)]/\delta x$$

For convenience, the function A(x,y) is used to describe the factor:

$$\frac{1}{1 - \alpha U(x,y)}$$

Thus, the term $\frac{\partial}{\partial x}\left[\frac{D_a}{1 - \alpha u}\frac{\partial u}{\partial x}\right]$ becomes:

$$\frac{\partial}{\partial x}\left[D_a A(x,y)\left(\frac{U(x + \delta x/2) - U(x - \delta x/2)}{\delta x}\right)\right]$$

Using a shorthand notation; e.g. $U(x+\frac{1}{2}\delta x,y) \rightarrow U_x^{\frac{1}{2}+}$ this becomes:

$$\frac{D_a}{\delta x^2}\left[A_x^{\frac{1}{2}+}.\left(U_x^{1+}-U\right)-A_x^{\frac{1}{2}-}.\left(U-U_x^{1-}\right)\right] \qquad \text{(A3.1.2)}$$

The term $A_x^{\frac{1}{2}+}$ is evaluated as the average of A and A_x^{1+}.

i.e. $$A_x^{\frac{1}{2}+} = [A_x^{1+} + A]/2$$

It can be shown[38] that a numerical algorithm based on A3.1.2 will be unconditionally unstable. This can be rectified by replacing U(x,y) by the time-average of NU(x,y) and LU(x,y).

i.e. $$U(x,y) \longrightarrow [NU(x,y) + LU(x,y)]/2$$

A3.1.2 becomes:

$$\frac{D_a}{\delta x^2}\left[A_x^{\frac{1}{2}+}U_x^{1+} + A_x^{\frac{1}{2}-}U_x^{1-} - \frac{1}{2}[A_x^{\frac{1}{2}+} + A_x^{\frac{1}{2}-}].[NU + LU]\right]$$

The y-direction term of 5.1.1, $\frac{\partial}{\partial y}\left[\frac{D_a}{1-\alpha u}\frac{\partial u}{\partial y}\right]$, reduces to a similar expression. Substituting these expressions into 5.1.1 (including A3.1.1), setting $\delta x = \delta y$ and rearranging to obtain an expression which is explicit for NU(x,y), finally yields:

$$NU(x,y) = \frac{[2r\Sigma AU + (1 - r\Sigma A)LU(x,y)]}{(1 + r\Sigma A)} \qquad \text{(A3.1.3)}$$

where $r = \frac{D_a \delta t}{\delta x \delta y}$

$$\Sigma AU = A_x^{\frac{1}{2}+}U_x^{1+} + A_x^{\frac{1}{2}-}U_x^{1-} + A_y^{\frac{1}{2}+}U_y^{1+} + A_y^{\frac{1}{2}-}U_y^{1-}$$

and $$\Sigma A = A_x^{\frac{1}{2}+} + A_x^{\frac{1}{2}-} + A_y^{\frac{1}{2}+} + A_y^{\frac{1}{2}-}$$

In theory A3.1.3 is stable for all values of r[38] (r is the normalised time-step). However in practice the solution is found to

become inaccurate if it is too large.

A3.1.3 is the main algorithm used by the computer program DIF. However only one initial time-level is available (the t = o boundary condition) so a 2-level algorithm is used as a starting routine to generate U and LU.

A3.2 The Two Level Method

Because of the instability of A3.1.2 a different approach must be used if only 2 time-levels are available. Expanding the term $\frac{\partial}{\partial x}\left[\frac{D_a}{1-\alpha u}\frac{\partial u}{\partial x}\right]$ we obtain:

$$\frac{D_a}{1-\alpha u}\left[\frac{\alpha}{1-\alpha u}\left(\frac{\partial u}{\partial x}\right)^2 + \frac{\partial^2 u}{\partial x^2}\right]$$

which is, in terms of finite differences:

$$D_a \cdot A\left[\alpha A\left(\frac{U_x^{1+} - U_x^{1-}}{2\delta_x}\right) + \left(\frac{U_x^{1+} + U_x^{1-} - 2U}{\delta_x^2}\right)\right]$$

Expanding the y-derivative similarly and using the first-order correct time derivative:

i.e. $$\frac{\partial u}{\partial t} \longrightarrow \frac{NU(x,y) - U(x,y)}{\delta t}$$

we obtain from 5.1.1:

$$NU(x,y) = \frac{r}{1-\alpha u}\left[\frac{\alpha}{4}\cdot\frac{1}{1-\alpha u}\left(\Delta_x U^2 + \Delta_y U^2\right) + \Sigma U - 4U\right] + U(x,y) \quad \text{(A3.2.1)}$$

where $U = U(x,y)$

$$\Sigma U = U_x^{1+} + U_x^{1-} + U_y^{1+} + U_y^{1-}$$

$$\Delta_x U = U_x^{1+} - U_x^{1-}$$

Stability

A3.2.1, like any other explicit 2-level algorithm, is stable only if r is sufficiently small. In this case it is possible to obtain the correct stability criterion, without entering into the complexities of numerical methods theory, by means of a fairly simple argument.

Instability of a numerical algorithm is caused by the build-up of errors over successive iterations. The gain factor which causes error magnification in A3.2.1, is a function of $\frac{r}{1 - \alpha u}$. This is maximised when $U \simeq 1$, so this is the most critical case. Also, to consider the worst possible case, the gain factor must remain high so $NU \simeq 1$. In this case $\Delta_x U$ and $\Delta_y U$ are negligible and $\Sigma U \simeq 4U$.

The worst possible error distribution is also the most likely. Finite difference solutions tend to oscillate slightly about the true solution, so if U is in error by $-\varepsilon$, each of its surrounding points is considered to be in error by $+\varepsilon$. (If this is so, then the ΔU terms are accurate and are not considered).

For the maximum possible error in NU;

$$NU \simeq \frac{r}{1 - \alpha u}\left[4U + 4\varepsilon \;\; - 4U + 4\varepsilon\right] + (U - \varepsilon)$$

So NU is in error by $2\frac{4r}{1 - \alpha u}\varepsilon - \varepsilon \quad (= \varepsilon' \text{ say})$

If $\frac{4r}{1 - \alpha u} = 1$, $\varepsilon' = -\varepsilon$ and the error magnitude remains unaltered.

If $\frac{4r}{1 - \alpha u} < 1$, error is attenuated.

If $\frac{4r}{1 - \alpha u} > 1$, error is amplified.

Thus, finally, the stability criterion for this algorithm is:

$$r < (1 - \alpha)/4 \qquad \text{(A3.2.2)}$$

A3.3 Implementation

The boundary condition code for each point (x,y) is stored in a fourth rectangular array NB(x,y). The various codes are explained in the main text. This integer number, NB(x,y) is used to dictate the precise form of A3.1.3 or A3.2.1 used to compute NU(x,y). In two cases (boundary codes 1 and 2) NU(x,y) is simply set to a constant.

The Program

After NB has been constructed from input data and the initial concentration profile (if any) read in from a data file the diffusion proceeds.

The 2-level starting routine, using a small value of r for reasons of stability, calculates the first few time levels after which, without changing r, the program moves on to the 3-level routine. After this, r is doubled every 8 time steps until the limiting value $r = r_0$ is reached. At each doubling of r, LU is left at its previous level to keep the 3-level computation accurate. This progression, shown diagramatically in Fig. A3.1, provides a smooth transition from a very small r ($r = r_0/64$ in practice) to the large, efficient value possible with the 3-level routine. A large r near the time origin is undesirable.

Further programming details are best dealt with by means of flow charts and require no further explanation.

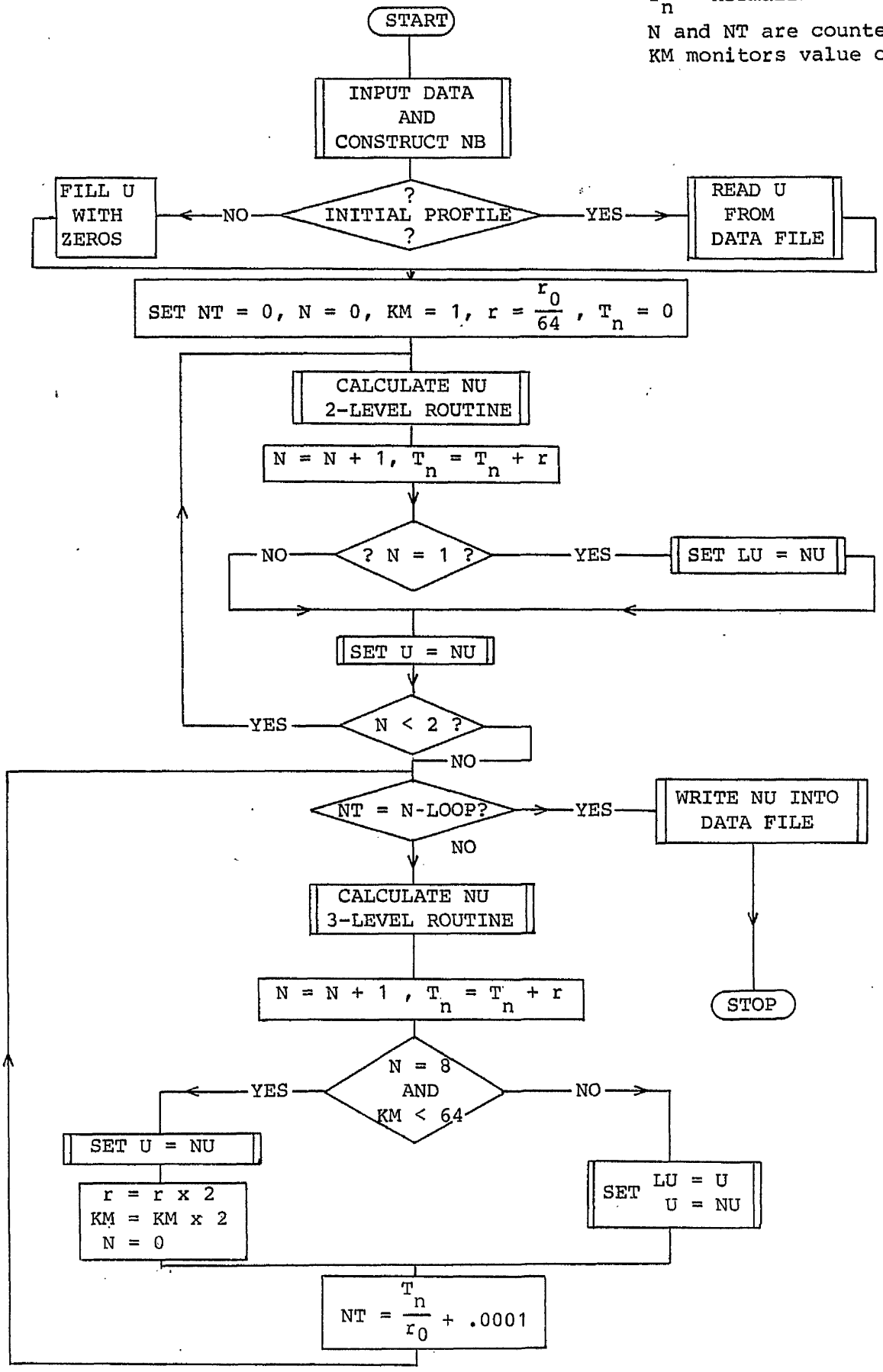
Flow Chart for DIF Main Program

Key

T_n = normalised Time

N and NT are counters

KM monitors value of r



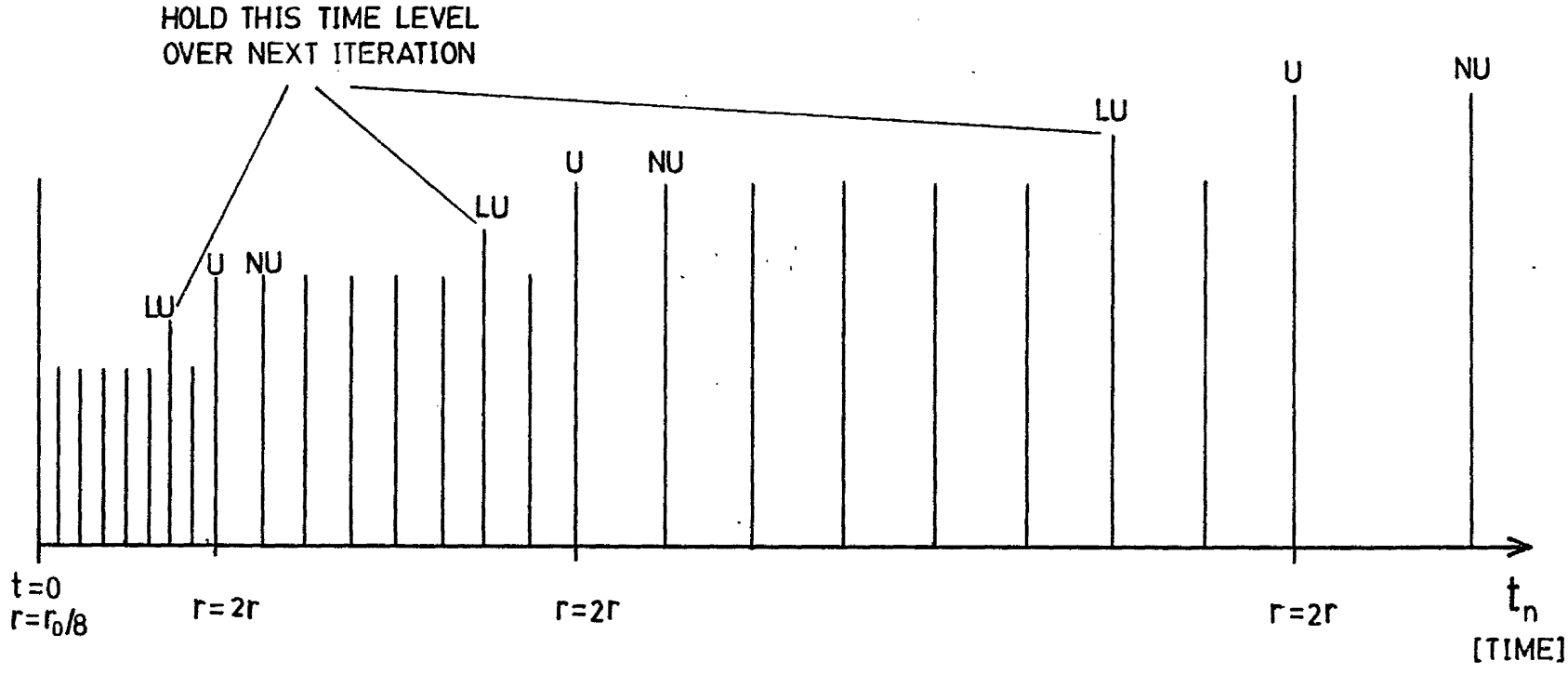


Fig. A3·1 Diagramatic representation of speed increase in DIF
[vertical lines are arrays of u values]

APPENDIX 4

The Variational Method

5.2.1 has the general form of an eigenvalue equation:

$$L(u) = \lambda u \qquad \text{(A4.1.1)}$$

where λ (or β^2), the eigenvalue, is constant for all x and y;

u (or E_x) is a function of x and y which satisfies the equation;

L (or $\nabla^2_{xy} + k^2$) is an operator which is self-adjoint and positive-definite.[60]

In order to obtain the functional A4.1.3 (below) A4.1.1 is multiplied by u. The condition that A4.1.1 be true for all x,y is expressed by integrating over all x,y:

So $$\int_{-\infty}^{\infty}\int L(u).u\,dxdy = \lambda\int_{-\infty}^{\infty}\int u.u\,dxdy \qquad \text{(A4.1.2)}$$

or $$\lambda = \frac{\langle L(u)\ ;\ u\rangle}{\langle u\ ;\ u\rangle} \qquad \text{(A4.1.3)}$$

where $\langle f\ ;\ g\rangle$ denotes the inner product of real functions f and g

and $$\langle f\ ;\ g\rangle = \int_{-\infty}^{\infty}\int f.g\,dxdy$$

In terms of a general function, $g(x,y)$, and a constant, γ, A4.1.3 becomes:

$$\gamma = \frac{\langle L(g)\ ;\ g\rangle}{\langle g\ ;\ g\rangle} \qquad \text{(A4.1.4)}$$

Let $u(x,y)$ and λ be particular examples of $g(x,y)$ and γ which satisfy the wave equation (A4.1.1). It can be shown that, if $g(x,y) = u(x,y)$ then (clearly) $\gamma = \lambda$ and:

$$\frac{\partial \gamma}{\partial \alpha} = 0$$

where α is some parameter of $g(x,y)$ which is varied to modify the function. This can be demonstrated by differentiating A4.1.4, setting to zero and noting the conditions for which the resultant equation is true. We obtain:

$$\langle g \; ; \; g\rangle\left[\langle g \; ; \; \frac{\partial}{\partial \alpha}\,[L(g)]\rangle + \langle L(g) \; ; \; \frac{\partial g}{\partial \alpha}\rangle\right] - 2\langle L(g) \; ; \; g\rangle\cdot\langle g \; ; \; \frac{\partial g}{\partial \alpha}\rangle = 0$$

Since L is self adjoint, and the order of operators L and $\frac{\partial}{\partial \alpha}$ can be reversed:

$$\langle g \; ; \; \frac{\partial}{\partial \alpha}\,[L(g)]\rangle = \langle L(g) \; ; \; \frac{\partial g}{\partial \alpha}\rangle$$

Since, according to A4.1.4

$$\langle L(g) \; ; \; g\rangle = \gamma\langle g \; ; \; g\rangle$$

we obtain

$$\langle L(g) \; ; \; \frac{\partial g}{\partial \alpha}\rangle - \langle \gamma g \; ; \; \frac{\partial g}{\partial \alpha}\rangle = 0 \qquad \text{(A4.1.5)}$$

One solution of A4.1.5 is:

$$L(g) = \gamma g$$

which is a restatement of A4.1.1. Thus when the wave equation is satisfied, the value (γ) of the functional, A4.1.4, is at an extremium (λ). If, for example, this is a maximum we note that λ is the highest possible value of γ and that as $g(x,y)$ approaches

$u(x,y)$ γ approaches λ. It follows from this that, if $g(x,y)$ does not pass through $u(x,y)$ as the parameter α is varied, the extremium, $\gamma = \gamma_{max}$, yet gives the closest approach.

The basis of the variational method is to find the extremia of the functional A4.1.4. These always give the best approximation to the true solution within the capabilities of the trial function $g(x,y)$.

A4.2 The Rayleigh-Ritz Method

We use a form of trial function [$g(x,y)$] that is intrinsically capable of satisfying the wave equation. An infinite series of basis functions, $\varphi_{ij}(x,y)$, which form a complete, orthogonal set, with variable coefficients, a_{ij}, may fit almost any differentiable, piecewise continuous function within the domain over which the basis functions are defined.

The choice of basis function is rather limited. The waveguide field functions go to zero at $\pm \infty$, so sines and cosines (giving a Fourier series) cannot be used. The functions must go to zero at $\pm \infty$, be orthogonal and be continuous at the argument origin; there is little choice but to use the parabolic cylinder functions. These (when expanded) consist of a Hermite polynomial multiplied by a Gaussion, the nth order function being defined as:

$$D_n(x) = (-1)^n e^{x^2/4} \frac{d^n}{dx^n}\left[e^{-x^2/2}\right] \qquad \text{(A4.2.1)}$$

They are orthogonal:

i.e. $$\left\langle D_i\left(\frac{x}{\xi}\right) ; D_k\left(\frac{x}{\xi}\right)\right\rangle = \begin{cases} 0 : i \neq k \\ \sqrt{2\pi}\, i!\xi : i = k \end{cases}$$

where ξ is a scaling factor.

To model a 2-D field-function, a series of 2-D basis functions is required. The product of x-direction and y-direction parabolic-cylinder functions is used.

Normalised and scaled:

$$\varphi_{ij}(x,y) = (2\pi\xi\eta i!j!)^{-\frac{1}{2}} D_i\left(\frac{x}{\xi}\right) D_j\left(\frac{y - y_0}{\eta}\right)$$

This is <u>orthonormal</u>

i.e.

$$\langle \varphi_{ij} ; \varphi_{kl} \rangle = \delta_{ik} \cdot \delta_{jl}$$

using the Kronecker-δ symbol:

$$\delta_{ik} = \begin{cases} 1 ; i = k \\ 0 ; i \neq k \end{cases}$$

Finally, the series is:

$$\psi(x,y) = \sum_{i=0}^{n} \sum_{j=0}^{m} a_{ij} (2\pi\xi\eta i!j!)^{-\frac{1}{2}} D_i\left(\frac{x}{\xi}\right) D_j\left(\frac{y - y_0}{\eta}\right) \qquad \text{(A4.2.2)}$$

The coefficients a_{ij} are varied until $\psi(x,y) = E_x(x,y)$ as defined by stationary values of the variational expression A4.1.3.

i.e. $$F = \frac{-\iint_{-\infty}^{\infty} (\nabla^2\psi + k^2(x,y)\psi) \cdot \psi \, dx\,dy}{\iint_{-\infty}^{\infty} \psi \cdot \psi \, dx\,dy} \qquad \text{(A4.2.3)}$$

For example, a true minimum is obtained when any small variation of any a_{ij} serves only to increase the value of F. Therefore we require $\frac{\partial F}{\partial a_{ij}} = 0$ simultaneously for all i and j.

(Note that because the product $\psi.\psi$ occurs, two distinct sets of counters ij and kl are required.) Thus, if pq is some particular value of ij or kl we require:

$$\frac{\partial}{\partial a_{pq}} \text{Num.} \; - F. \frac{\partial}{\partial a_{pq}} \text{Den.} \; = 0 \qquad \text{(A4.2.3)}$$

simultaneously for all pq (Num. and Den. are the numerator and denominator of functional A4.2.3).

In both numerator and denominator, the product of sums is re-written as a sum of products and the order of summation and integration is reversed yielding:

$$\text{Num.} = \sum_{ij=0}^{mn} \sum_{kl=0}^{mn} N_{ijkl} \cdot a_{ij} \cdot a_{kl}$$

$$\text{Den.} = \sum_{ij=0}^{mn} \sum_{kl=0}^{mn} D_{ijkl} \cdot a_{ij} \cdot a_{kl} \qquad \text{(A4.2.4)}$$

where $D_{ijkl} = \delta_{ik} \cdot \delta_{jl}$

N_{ijkl} is rather complicated (see section A4.3) but is symmetrical, i.e. $N_{pqrs} = N_{rspq}$.

Expanding A4.2.4 (assuming symmetry)

$$\text{Num.} = N_{pqpq} a_{pq}^2 + 2. \sum_{\substack{ij \neq pq \\ 0}}^{mn} N_{ijpq} a_{ij} a_{pq} + \sum_{\substack{ij \neq pq \\ 0}}^{mn} \sum_{\substack{kl \neq pq \\ 0}}^{mn} N_{ijkl} a_{ij} a_{kl}$$

Thus: $$\frac{\partial}{\partial a_{pq}} (\text{Num.}) = 2 \sum_{ij=0}^{mn} N_{ijpq} a_{ij}$$

Similarly: $$\frac{\partial}{\partial a_{pq}} (\text{Den.}) = 2 \sum_{ij=0}^{mn} D_{ijpq} a_{ij}$$

A4.2.3 becomes:

$$\sum_{ij=0}^{mn} [N_{ijkl} - F\delta_{ik}\delta_{jl}]a_{ij} = 0 \quad (kl \text{ fixed})$$

A similar equation is set up for each kl. When this system of equations is complete F may be replaced by β^2.

In matrix form $$[\underset{\sim}{N} - \beta^2\underset{\sim}{I}]\underset{\sim}{a} = \underset{\sim}{0} \qquad \text{(A4.3.5)}$$

where $\underset{\sim}{N}$ is a square, symmetric matrix (mn x mn) whose elements are N_{ijkl}, having eigenvalues β^2 and eigen-vectors $\underset{\sim}{a}$ ($\underset{\sim}{a}$ is a column matrix).

A4.3 The Form of N_{ijkl}

The numerator of A4.2.3 is reduced to the form:

$$\sum_{ij=0}^{mn}\sum_{kl=0}^{mn}\iint_{-\infty}^{\infty}\frac{a_{ij}a_{kl}}{2\pi\xi\eta\sqrt{i!j!k!l!}}D_i\left(\frac{x}{\xi}\right)D_j\left(\frac{y-y_0}{\eta}\right)\left[(\nabla^2+k^2)D_k\left(\frac{x}{\xi}\right)D_l\left(\frac{y-y_0}{\eta}\right)\right]dxdy$$

Thus N_{ijkl} is the sum of three parts:

$$\frac{k_0^2}{2\pi\xi\eta\sqrt{i!j!k!l!}}\iint_{-\infty}^{\infty}D_i\left(\frac{x}{\xi}\right)D_j\left(\frac{y-y_0}{\eta}\right)D_k\left(\frac{x}{\xi}\right)D_l\left(\frac{y-y_0}{\eta}\right)n^2(x,y)\,dxdy \qquad \text{(i)}$$

$$+\ \frac{1}{2\pi\xi\eta\sqrt{i!j!k!l!}}\iint_{-\infty}^{\infty}D_i\left(\frac{x}{\xi}\right)D_j\left(\frac{y-y_0}{\eta}\right)\left[\frac{\partial^2}{\partial x^2}\left[D_k\left(\frac{x}{\xi}\right)D_l\left(\frac{y-y_0}{\eta}\right)\right]\right]dxdy \qquad \text{(ii)}$$

$$+\ \frac{1}{2\pi\xi\eta\sqrt{i!j!k!l!}}\iint_{-\infty}^{\infty}D_i\left(\frac{x}{\xi}\right)D_j\left(\frac{y-y_0}{\eta}\right)\left[\frac{\partial^2}{\partial y^2}\left[D_k\left(\frac{x}{\xi}\right)D_l\left(\frac{y-y_0}{\eta}\right)\right]\right]dxdy \qquad \text{(iii)}$$

(A4.3.1)

Because of the n(x,y) factor, the first part cannot be further modified but is calculated by numerical integration.

The second and third parts may be separated into x and y integrals. Part (ii), for example, becomes:

$$\frac{1}{2\pi\xi\eta\sqrt{i!j!k!l!}}\int_{-\infty}^{\infty} D_j\left(\frac{y-y_0}{\eta}\right) D_l\left(\frac{y-y_0}{\eta}\right)dy \cdot \int_{-\infty}^{\infty} D_i\left(\frac{x}{\xi}\right)\frac{\partial^2}{\partial x^2} D_k\left(\frac{x}{\xi}\right) dx$$

$$= \frac{\delta_{jl}}{\xi\sqrt{2\pi i!k!}}\int_{-\infty}^{\infty} D_i\left(\frac{x}{\xi}\right)\frac{\partial^2}{\partial x^2} D_k\left(\frac{x}{\xi}\right) dx \qquad \text{(A4.3.2)}$$

Using the definition of $D_i\left(\frac{x}{\xi}\right)$ and some fairly subtle integration (by parts) this can be reduced to:

$$\delta_{jl}\cdot\frac{1}{\xi^2}\left[\frac{1}{4}\sqrt{k(k-1)}\,\delta_{i,k-2} - \left(\frac{1}{4}+\frac{1}{2}k\right)\delta_{ik} + \frac{1}{4}\sqrt{(k+1)(k+2)}\,\delta_{i,k+2}\right] \qquad \text{(A4.3.3)}$$

Part (iii) is, of course, exactly similar.

A4.4 Practical Details

The region in which the wave equation is to be solved, illustrated in Fig. A4.1, is assumed to be symmetrical about the y-axis.

The second and third parts of N_{ijkl} are calculated from A4.3.3 but the first part (see A4.3.1) is more complicated. The integral is performed in four stages:

1) The lower half plane ($n = n_a$)

2) The entire upper half plane ($n = n_s$)

3) The integral over the rectangular area ($-x_1 < x < x_1$; $0 < y < y_1$) for $n = n_s$ is subtracted

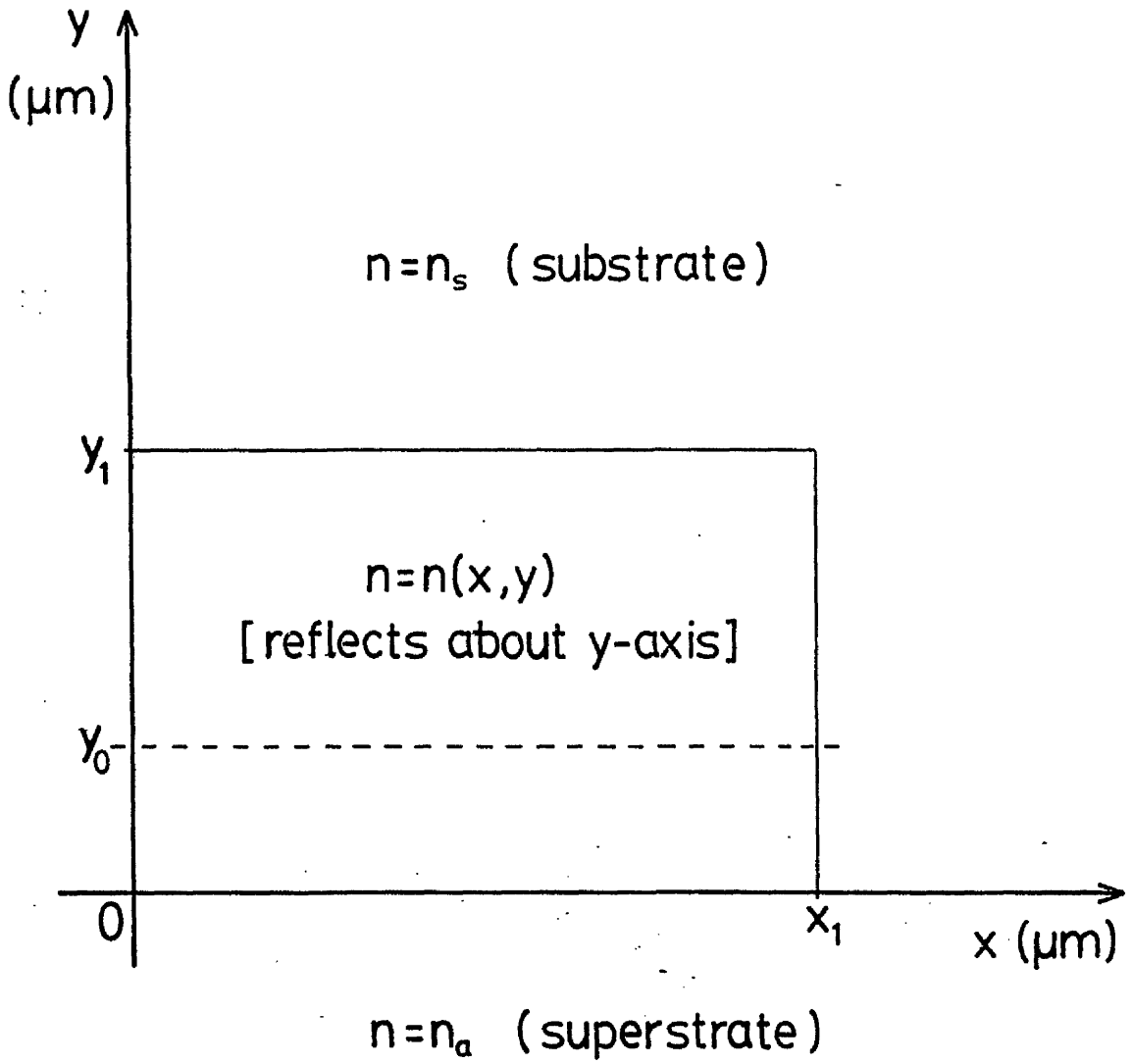


Fig. A4·1 Co-ordinate and refractive index assumptions pertaining to WAVE.

4) The integral over the above rectangular area for n = n(x,y) is added.

The 2-D integral required when n = n(x,y) is performed using Simpson's rule in two dimensions. When the refractive index is constant, the integrals may be broken down into integrals of form:

$$\frac{1}{\eta}\int_{-z_0}^{z_1-z_0} D_j\left(\frac{u}{\eta}\right) \cdot D_1\left(\frac{u}{\eta}\right) du$$

(which is calculated by Simpson's rule)

and:
$$\frac{1}{\eta}\int_{0}^{\infty} D_j\left(\frac{u}{\eta}\right) D_1\left(\frac{u}{\eta}\right) du \qquad \text{(A4.4.1)}$$

which, being dependent only upon j and 1 is stored in an array.

The Semi-infinite Integral

If j = 1 then integral A4.4.1 is simply half of the infinite integral (since the product of equal functions is symmetric and always positive).

$$I_{j1} = \sqrt{\frac{\pi}{2}}\ j! \quad : \quad j = 1$$

If $j \neq 1$ integration by parts yields the following formula:

$$I_{j1} = \sum_{k=0}^{1-1} \left[(-1)^{(1+j+1-k)} \cdot C_0(1-k-1) \cdot C_k(j) \cdot k! \right]$$

Where $C_0(1-k-1)$ is the constant term of the $(1-k-1)$th order, Hermite-type polynomial:

$$D_{1-k-1}(x)e^{x^2/4}$$

Similarly, $C_k(j)$ is the k th order coefficient of a j th order polynomial.

It is far simpler to derive the array of values according to the following empirical rule which has been checked, from the above, for up to D_{20}.

Working by diagonals:

1) If j and 1 are both even or both odd the diagonal is zero. An exception is the leading diagonal as given above (j = 1).

2) The first column (j = 0) is given by a "factorial" of negative, odd numbers alternating with zeros:

$$\left[\sqrt{\frac{\pi}{2}}\right], 1,\ 0,\ -1,\ 0,\ 3,\ 0,\ -15,\ 0,\ 105,\ 0,\ -945,\ \ldots$$

3) On each non-zero diagonal (except j = 1), each element is given by the previous element (j - 1, 1 - 1) multiplied by an integer factor. This factor is not constant but varies according to the sequence:

$$1,\ 3 + m,\ 3,\ 5 + m,\ 5,\ 7 + m,\ 7,\ 9 + m,\ 9 \ldots .$$

where $m = |1 - j| - 1$ is constant for each diagonal.

In the program, this rule is used to generate the array of integral values automatically.

Evaluation of the Parabolic-Cylinder Function

The coefficients of the Hermite-type polynomial:

$$D_n(x)e^{x^2/4} = 2^{-n/2}H_n\left[\frac{x}{\sqrt{2}}\right]$$

are calculated by the formula:

$$C_m = -C_{m+2} \cdot \frac{(m+2)!}{m!\,(n-m)}$$

where n is the order of the polynomial and C_n is always 1 (even

order) or -1 (odd order). $C_{m\ \pm 1}$ is zero (the function is always either symmetric or antisymmetric).

An array of these coefficients is generated automatically.

The Program

The parameters η, ξ and y_0 are selected by maximising N_{0000} - i.e. by finding the parameters of the best-fitting 2-D gaussian function. A separate program is used for this. N.A.G. (Numerical Algorithms Group) routines are used for this maximisation and also to derive the eigenvalues and eigenvectors from the matrix $\underset{\sim}{N}$.

The programming is straightforward, consisting of four nested loops to work through the counters i, j, k, l computing N_{ijkl} as described. Only half of the symmetric matrix $\underset{\sim}{N}$ is computed and the assumed lateral symmetry allows exclusively even or odd order x-direction basis functions to be used, giving exclusively even or odd order lateral modes.

APPENDIX 5

An Alternative Technique for the Fabrication of Anodised Diffusion Masks

The technique to be described was designed to permit the anodisation of isolated areas of a waveguide pattern. It is based on the lift-off techniques described in section 6.1.

After cleaning the slide, a thin film (~ 200 Å) of aluminium is deposited over the entire surface. The subsequent steps follow the usual pattern of photo-resist deposition, exposure, development and a second, thicker (~1000 Å) aluminisation. At this stage, the only difference between this and the usual situation, immediately before lift-off, is the presence of the initial, thin metal layer which electrically connects all areas of the pattern.

The anodisation is carried out immediately prior to lift-off. The thin, connecting film is protected by the resist ridges and, after lift-off, the result is an anodised mask whose diffusion apertures are partially filled with metallic aluminium. This is removed by etching. If the etch is cold (< 20°C) it attacks the Al_2O_3 very much more slowly than the aluminium and it is found that good masks can result. This was not always the case, however, and the technique (like standard anodisation) requires further investigation.

REFERENCES

(1) Unger, H.G., "Planar Optical Waveguides and Fibres", Oxford, Clarendon Press, 1977.

(2) Marcuse, D., "Theory of Dielectric Optical Waveguides", London, Academic Press, 1974.

(3) Marcuse, D., "Light Transmission Optics", New York, Van Nostrand, 1972.

(4) Arnaud, J.A., "Beam and Fibre Optics", London, Academic Press, 1976.

(5) Bremner, H., "The W.K.B. Approximation as the first term of a Geometric Optical Series", Commun. Pure Appl. Math., Vol. 4, p. 105, 1951.

(6) Dicke, R.H., Wittke, J.P., "Introduction to Quantum Mechanics", Addison-Wesley, 1960.

(7) Gedeon, A., "Comparison between Rigorous Theory and W.K.B.-Analysis of modes in Graded Index Waveguides", Optics Commun., 12, pp. 329-332, 1974.

(8) Brekhovskikh, L.M., "Waves in Layered Media", New York, Academic Press, 1960.

(9) Marcatili, E.A.J., "Dielectric Rectangular Waveguide and Directional Coupler for Integrated Optics", Bell Syst. Tech. Journal, 48, pp. 2071-2102, 1969.

(10) Goell, J.E., "A Circular-Harmonic Computer Analysis of Rectangular Dielectric Waveguides", Bell Syst. Tech. Journal, 48, pp. 2133-2160, 1969.

(11) Hocker, G.B., Burns, W.K., "Mode Dispersion in Diffused Channel Waveguides by the Effective Index Method", Opt. Commun., Vol. 12, No. 3, p. 329, 1974.

(12) Knox, R.M., Toulios, P.P., "Proceedings of MRI Symposium on Submillimeter Waves", (Ed. J. Fox), Brooklyn, Polytechnic Press, 1970.

(13) Snyder, A.W., Young, W.R., "Modes of Optical Waveguides", J. Opt. Soc. Am., Vol. 68, No. 3, p. 297, March 1978.

(14) Ikuno, H., "Analysis of Wave Propagation in Inhomogeneous Dielectric Slab Waveguides", I.E.E.E. Trans. on Microwave Theory and Techniques, 26, No. 4, p. 261, April 1978.

(15) Cullen, A.L., "Coupled, Parallel, Rectangular, Dielectric Waveguides", Proc. I.E.E., Vol. 122, No. 6, p. 593, June 1975.

(16) Findakly, T., Chin-Lin, C., "Optical Directional Couplers with Variable Spacing", App. Opt., Vol. 17, No. 5, p. 769, March 1978.

(17) Ankiewicz, A., Pask, C., "Tunneling Rays in Graded-Index Fibres", Opt. and Quant. Electron., 10, pp. 83-93, 1978.

(18) Love, J.D., Winkler, C., "A Universal Tunneling Coefficient for Step and Graded-Index Multimode Fibres", Opt. and Quant. Electron., 10, pp. 341-351, 1978.

(19) Ramaswamy, V., Standley, R.D., "Radiation Patterns from Parallel, Optical Waveguide Directional Couplers - Parameter Measurements", Bell Syst. Tech. Journal, Vol. 57, No. 9, p. 2685, Sept. 1978.

(20) Itanami, T., Shindo, S., "Channel Dropping Filter for Millimeter-Wave Integrated Circuits", I.E.E.E. Trans. on Microwave Theory and Techniques", Vol. MTT-26, No. 10, p. 759, Oct. 1978.

(21) Shindo, S., Itanami, T., "Low Loss Rectangular Dielectric Image Line for Millimeter-Wave Integrated Circuits", I.E.E.E. Trans on Microwave Theory and Techniques, Vol. MTT-26, No. 10, p. 747, Oct. 1978.

(22) Miyagi, M., Gar Lam Yip, "Design Theory of High-Q Optical Ring Resonator with asymmetric Three Layered Dielectrics", Opt. and Quant. Electron., 10, pp. 425-433, 1978.

(23) Taylor, H.F., "Power Loss at Directional Change in Dielectric Waveguides", Applied Optics, Vol. 13, No. 3, p. 642, March 1974.

(24) Neumann, E.G., Rudolf, H.D., "Radiation from Bends in Dielectric Rod Transmission Lines", I.E.E.E. Trans. on Microwave Theory and Techniques, Vol. MTT-23, No. 1, p. 142, Jan. 1975.

(25) Gambling, W.A. et al., "Measurement of Radiation Loss in Curved Single Mode Fibres", IEE Journal on Microwaves, Optics and Acoustics, Vol. 2, No. 4, p. 134, July 1978.

(26) Marcatili, E.A.J., "Bends in Optical Dielectric Waveguides", Bell Syst. Tech. Journal, 48, p. 2103, Sept. 1969.

(27) Marcatili, E.A.J., Miller, S.E., "Improved Relations Describing Directional Control in Electromagnetic Wave Guidance", Bell Syst. Tech. Journal, 48, p. 2161, Sept. 1969.

(28) Snyder, A.W., Love, J.D., "Reflection at a Curved Dielectric Interface - Electromagnetic Tunneling", I.E.E.E. Trans. on Microwave Theory and Techniques, Vol. MTT-23, No. 1, p. 134, Jan. 1975.

(29) Snyder, A.W., White I., Mitchell, D.J., "Radiation from Bent Optical Waveguides", Electronics Letters, Vol. 11, No. 15, p. 332, 24 July 1975.

(30) White, I.A., "Radiation from Bends in Optical Waveguides: The Volume Current Method", I.E.E. Journal on Microw. Opt. Acoust., Vol. 3, No. 5, p. 186, Sept. 1979.

(31) Heiblum, M., Harris, J.H., "Analysis of Curved Optical Waveguides by Conformal Transformation", I.E.E.E. Journal of Quant. Elect., Vol. Q-E 11, No. 2, p. 75, Feb. 1975.

(32) Heiblum, M., "The Prism Coupler and the Dielectric Bend: Similarities and Anomalous Behaviour", I.E.E.E. Journal of Quant. Elect., Vol. Q-E 12, No. 8, p. 463, Aug. 1976.

(33) Gedeon, A. Carnstam, B., "Radiation Loss in Bends of Diffused Optical Stripe Guides", Opt. and Quant. Elect., 7, p. 456, 1975.

(34) Harris, J.H., Schubert, R., Polky, J.N., "Beam Coupling to Films", J. Opt. Soc. Am., Vol. 60, No. 8, p. 1007, Aug. 1970.

(35) Crank, J., "The Mathematics of Diffusion", London, Oxford University Press, 1975.

(36) Ames, W.F., "Numerical Methods for Partial Differential Equations", London, Academic Press/Nelson, 1977.

(37) Von Rosenberg, D.U., "Methods for the Numerical Solution of Partial Differential Equations", New York, Elsevier, 1969.

(38) Mitchel, A.R., "Computational Methods in Partial Differential Equations", London, Wiley, 1969.

(39) Carrier, G.F., Pearson, C.E., "Partial Differential Equations - Theory and Technique", London, Academic Press, 1976.

(40) Doremus, R.H., "Ion Exchange in Glasses", Chapter 1 of "Ion Exchange - a Series of Advances", Ed. by Marinsky, J.A., New York, Dekker, 1969.

(41) Holland, L., "The Properties of Glass Surfaces", London, Chapman and Hall, 1964.

(42) Malm, D.L. et al., "Depth Profiles of Sodium and Calcium in Glasses", J. Vacu. Sci. Technol., Vol. 15, No. 1, Jan./Feb. 1978.

(43) Giallorenzi et al., "Optical Waveguides Formed by Thermal Migration of Ions in Glass", Applied Optics, Vol. 2, No. 6, p. 1240, June 1973.

(44) Stewart, G. et al., "Planar Optical Waveguides Formed by Silver-Ion Migration in Glass", I.E.E.E. J. of Quantum Elect., Vol. QE-13, No. 4, p. 192, April 1977.

(45) Millar, C.A., Hutchins, R.H., "Manufacturing Tolerances for Silver-Sodium Ion-Exchanged Planar Optical Waveguides", J. Phys. D: Appl. Phys., Vol. 11, p. 1567, 1978.

(46) Chartier, G.H. et al., "Optical Waveguides Fabricated by Electric-Field Controlled Ion-Exchange in Glass", Electron. Letters, Vol. 14, p. 132, 1978.

(47) Chartier, G.H. et al., "Graded Index Surface or Buried Waveguides by Ion-Exchange in Glass", unpublished, 1979.

(48) Wilkinson, C.D.W., Walker, R., "The Diffusion Profile of Stripe Optical Waveguides formed by Ion Exchange", Electron. Letters, Vol. 14, No. 18, p. 599, Aug. 1978.

(49) Stewart, G., Laybourn, P.J.R., "Fabrication of Ion-Exchanged Optical Waveguides from Dilute Silver Nitrate Melts", I.E.E.E. J. of Quant. Electron., Vol. QE-14, No. 12, p. 930, Dec. 1978.

(50) Doremus, R.H., "Exchange and Diffusion of Ions in Glass", J. of Phys. Chem., Vol. 68, Vol. 8, Aug. 1964.

(51) Taylor, H.F., "Dispersion Characteristics of Diffused Channel Waveguides", I.E.E.E. J. of Quant. Electron., Vol. QE-12, No. 12, p. 748, Dec. 1976.

(52) Matsuhara, M., "Analysis of T.E.M. Modes in Dielectric Waveguides by a Variational Method", J. Opt. Soc. Am., Vol. 63, No. 12, p. 1514, Dec. 1973.

(53) Jeffreys, H and B.S., "Methods of Mathematical Physics", Cambridge, C.U.P., 1956.

(54) Pelosi, P.M. et al., "Propagation Characteristics of Trapezoidal Cross-Section Ridge Optical Waveguides: an Experimental and Theoretical Investigation", Appl. Opt., Vol. 17, p. 1187, April 1978.

(55) Gallagher, J.G., "Silver Ion-Exchanged Waveguides for Integrated Optics", PhD. Thesis, University of Glasgow, 1977.

(56) Abramowitz, M., Stegun, I.A., "Handbook of Mathematical Functions", New York, Dover, 1965.

(57) Swyt, D.A., "N.B.S. Program in Photomask Linewidth Measurements", Solid State Technology, p. 55, April 1976.

(58) Wernick, S., Pinner R., "Surface Treatment of Aluminium", (Two Volumes), Fourth Edition, Teddington, Robert Draper Ltd., 1972.

(59) Geshiro, M. et al., "Modal Analysis of Strip-Loaded Diffused Optical Waveguides by a Variational Method", I.E.E.E. J. of Quant. Electron., Vol. QE-14, No. 4, p. 259, April 1978.

(60) Wexler, A., "Computation of Electromagnetic Fields", I.E.E.E. Trans. on Microwave Theory and Techniques, MTT-17, No. 8, p. 416, Aug. 1969.

(61) Wilson, C.S., "Passive Waveguide Components for Integrated Optics", Ph.D. Thesis, Glasgow University, 1979.

(62) Resnick, R., Halliday, D., "Physics", New York, Wiley, 1966, p. 1099.

(63) Boltz, R.E., Tuve, G.L., "Handbook of Tables for Applied Engineering Science", Cleveland, Ohio, The Chemical Rubber Company, 1970.

(64) Yi-Yan, A., Wilkinson, J.A.H., Wilkinson, C.D.W., "Optical Waveguide Filters for the Visible Spectrum", I.E.E. Proc., Vol. 127, Pt. H, No. 6, p. 335, Dec. 1980.

(65) Haavisto, J., Pajer, G.A., "Characteristics of Thin Film Ring Resonators", submitted for presentation at the Third International Conference on Integrated Optics and Optical Fibre Communication (IOOC, 1981), unpublished.